

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>5</sup> :</b> <b>C07H 21/02, 21/04, C07K 15/00, C12Q 1/68, C12P 19/34</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 94/21663</b> <b>(43) International Publication Date:</b> 29 September 1994 (29.09.94)
<b>(21) International Application Number:</b> PCT/US94/03246 <b>(22) International Filing Date:</b> 24 March 1994 (24.03.94)  <b>(30) Priority Data:</b> 08/036,481 24 March 1993 (24.03.93) US 08/181,629 13 January 1994 (13.01.94) US  <b>(71) Applicant:</b> MOLECULAR BIOLOGY RESOURCES, INC. [US/US]; 5520 West Burleigh Street, Milwaukee, WI 53210 (US).  <b>(72) Inventors:</b> MEAD, David; 2301 Waltham Road, Madison, WI 53711 (US). SWAMINATHAN, Neela; 476 Presidential Lane, Madison, WI 53711 (US). VAN ETTEN, James; 7409 Stevens Ridge Road, Lincoln, NE 68516 (US). SKOWRON, Piotr; 7400 W. Center Street, Wauwatosa, WI 53210 (US).  <b>(74) Agent:</b> CLOUGH, David, W.; Marshall, O'Toole, Gerstein, Murray & Borun, 6300 Sears Tower, 223 S. Wacker Drive, Chicago, IL 60606-6402 (US).	<b>(81) Designated States:</b> AU, CA, FI, JP, SE, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>With international search report.</i>	
<b>(54) Title:</b> DINUCLEOTIDE RESTRICTION ENDONUCLEASE PREPARATIONS AND METHODS OF USE  <b>(57) Abstract</b>  The present invention is directed to materials and methods for the quasi-random and complete fragmentation of DNA using restriction endonuclease reagents capable of cutting DNA at a dinucleotide sequence. The invention is also directed to methods for labeling DNA, for shotgun cloning, for sequencing of DNA, for epitope mapping and for anonymous primer cloning, all using fragments of DNA generated by the method of the present invention. In addition, the present invention is directed to DNA sequences encoding a novel restriction endonuclease (designated R. Cvi II) and variants thereof as well as to methods and materials for production of the same by recombinant methods. A bacterial host cell transformed with DNA encoding R. Cvi II is also disclosed as well as methods for expressing R. Cvi II in the bacterial host system and subsequent materials and methods for purifying the enzyme.		

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

## DINUCLEOTIDE RESTRICTION ENDONUCLEASE PREPARATIONS AND METHODS OF USE.

### FIELD OF THE INVENTION

5           The present invention relates generally to isolated purified polynucleotides which encode restriction enzymes and to methods of expressing the restriction enzymes from such polynucleotides. More particularly this invention relates to isolated purified polynucleotides which encode CviII and related methods for the production of this enzyme.

10           Other aspects of the invention relate to methods for partially or completely digesting DNA at a dinucleotide sequence. More particularly, this aspect of the invention relates to methods of generating quasi-random fragments of DNA, and methods of cloning, labeling, and sequencing DNA, as well as epitope mapping of proteins. The invention also relates to methods for generating  
15           sequence-specific oligonucleotides from DNA, without prior knowledge of the nucleic acid sequence of such DNA, and to methods for cloning and labeling DNA after restriction digestion by a two base recognition endonuclease reagent. This invention also relates to methods for cloning, labeling, and detecting nucleic acids using two base restriction endonuclease reagents, such as CviI I, BsuR I,  
20           Aci I or CGase I. Further the invention relates to labeling DNA by taking advantage of certain properties of the holo-enzyme of thermostable DNA polymerases.

### BACKGROUND OF THE INVENTION

25           Restriction endonucleases are a group of enzymes originally found to be expressed in a wide variety of prokaryotic organisms. More recently they have also been found to be encoded in viral genomes. These enzymes catalyze the selective cleavage of DNA at generally short sequences, often unique to the individual enzyme. This ability to cleave makes restriction endonucleases indispensable tools in recombinant DNA technology. The increased commercial

- 2 -

availability of the isolated enzymes has contributed in large part to the enormous expansion in the field of recombinant DNA technology over the last few years.

These enzymes have been classified into three groups. Because of properties of the type I and type III enzymes, they have not been widely used in molecular biology applications, and will not be discussed further. Type II enzymes are part of a binary system known as a restriction modification system consisting of a restriction endonuclease that cleaves a specific sequence of nucleotides and a separate DNA modifying enzyme that modifies the same recognition sequence and thereby prevents cleavage by the cognate endonuclease. A total of about 2103 restriction enzymes are known, encompassing 179 different type II specificities (Roberts, *et al.*, *Nucl. Acids Res.* 20:2167-2180 (1992)). Although there are more than 1200 type II restriction enzymes, many of them are members of groups which recognize the same sequence. Restriction enzymes that recognize the same sequence are said to be isoschizomers.

The vast majority of type II restriction enzymes recognize specific double-stranded sequences which are four, five, or six nucleotides in length and which display twofold (palindromic) symmetry. A few enzymes recognize longer sequences or degenerate sequences.

The location of cleavage sites within a palindrome differs from enzyme to enzyme. Some enzymes cleave both strands exactly at the axis of symmetry generating fragments of DNA that carry blunt ends, while others cleave each strand at similar sequences on opposite sides of the axis of symmetry, creating fragments of DNA that carry protruding, single-stranded termini.

Restriction endonucleases with shorter recognition sequences cut DNA more frequently than those with longer recognition sequences. For example, assuming a 50% G-C content, a restriction endonuclease with a 4-base recognition sequence will cleave, on average, every  $4^4$  (256) bases compared to every  $4^6$  (4096) bases for a restriction endonuclease with a 6-base recognition sequence. Under certain conditions some restriction endonucleases are capable

- 3 -

of cleaving sequences which are similar but not identical to their defined recognition sequence. This altered specificity has been termed "star" (\*) activity and is observed only under certain non-standard reaction conditions. The manner in which an enzyme's specificity is altered depends on the particular enzyme and on the conditions employed to induce the star activity. Conditions that contribute to star activity include high glycerol concentration, high ratio of enzyme to DNA, low ionic strength, high pH, the presence of organic solvents, and the substitution of  $Mg^{++}$  with other divalent cations. The most common types of star activity involve cutting at a recognition sequence having a single base substitution, cutting at sites having truncation of the outer bases of the recognition sequence, and single-strand nicking. The following restriction endonucleases show star activity: Ase I, BamH I, BssH II, BsuR I, CviJ I, EcoR I, EcoR V, Hind III, Hinf I, Kpn I, Pst I, Pvu II, Sal I, Sca I, Taq I, and Xmn I. Star activity is generally viewed as undesirable, and of little intrinsic value.

Of the 179 unique type II restriction endonucleases, 31 have a 4-base recognition sequence, 11 have a 5-base recognition sequence, 127 have a 6-base recognition sequence, and 10 which have recognition sequences of greater than 6 bases. In two cases, a restriction endonuclease has a recognition sequence of less than 4 bases.

The restriction enzyme CviJ I has a three base recognition sequence or a two-base recognition sequence, depending on the reaction conditions. Under normal reaction conditions CviJ I recognizes the sequence PuGCPy (wherein Pu=purine and Py=pyrimidine) and cleaves between the G and C to leave blunt ends (Xia et al., 1987. *Nucleic Acids Res.* 15:6075-6090). Under "relaxed" or "star" conditions (in the presence of 1 mM ATP and 20 mM DTT) the specificity of CviJ I may be altered to cleave DNA more frequently. This activity is referred to as CviJ I\*, for star or altered specificity. However, CviJ I\* activity is not observed under conditions which favor star activity of other restriction endonucleases.

- 4 -

The restriction enzyme BsuR I normally recognizes the sequence GGCC and cleaves between the G and C to leave blunt ends. (Heininger, *et al.*, *Gene* 1:291-303 (1977)). Under relaxed conditions (high pH, low ionic strength, and high glycerol concentration) the specificity of Bsu RI may be altered to cleave DNA more frequently. An isoschizomer of this enzyme, Hae III, does not display this star activity.

In bacteria, the restriction endonuclease provides a mechanism of defense against foreign DNA molecules (e.g., bacteriophage DNA) by virtue of its ability to distinguish and cleave only exogenous DNA, leaving endogenous bacterial DNA unaffected. Viral endonucleases possess the same discerning capabilities, but rather than providing a means for defense, this activity has presumably evolved to cripple the host's ability to replicate its own DNA and allows the virus to assume control of the host's replication machinery.

Bacteria and viruses which express restriction endonucleases necessarily possess the inherent ability to protect their own genome from cleavage by their endogenous endonuclease. The primary mechanism by which this is accomplished is by modifying the organisms own DNA by, for example methylating a base in the recognition sequence which prevents binding and cleavage by the endonuclease. Therefore, to insure viability, the genome of an organism which expresses a restriction endonuclease is almost always heavily modified, usually by methylation of cytosine or adenosine bases. The methylase enzyme which modifies the genome (itself a useful tool in molecular biology) acts in tandem with the endonuclease, either as part of an enzyme complex (restriction/modification complex) or as two distinct entities. Therefore, recognizing that an organism expresses an enzyme with endonuclease activity strongly suggests the expression of an associated modifying methylase enzyme (and vice versa) and this association has led to isolation and cloning of a number of commercially available restriction/modification enzymes for use in the laboratory as discussed below.

- 5 -

One of the limitations in the use of restriction endonucleases exists when cleavage of a given sequence is required and no known endonuclease exists which is specific for that particular sequence. Therefore, the continued identification and isolation of unique restriction endonucleases and altered reaction conditions will allow for even more sophisticated manipulation of DNA *in vitro*.

A number of publications and patents describe the cloning of DNAs encoding restriction endonucleases. Included among these publications is Kiss, A., *et al.*, *Nucleic Acid Research* 13:6403-6421 (1985), which describes the cloned nucleotide sequence of the *BsuRI* restriction-modification system isolated from *Bacillus subtilis*. This system is specific for the sequence 5'-GGCC-3' and is defined by two gene products which are transcribed by different promoters. The methylase component of the system shows homology to the methylase from the *BspRI* and *SPR* restriction-modification systems.

Nwanko, D.O. and Wilson, G.G. *Gene* 64:1-8 (1988), describe the cloning and expression of the *MspI* restriction and modification genes isolated from *Moraxella* sp. This system recognizes the sequence 5'-CCGG-3' and both enzymes are functional in *E. coli*. Evidence indicates that these genes are transcribed in opposite directions, thus are probably under the control of different promoters.

Ashok, K.D., *et al.*, *Nucleic Acids Research* 20:1579-1585 (1992), describe the purification and characterization of cloned *MspI* methyltransferase, over-expressed in *E. coli*. At low concentrations the enzyme exists as a monomer, but at higher concentrations it exists mainly as a dimer. Polyclonal antibodies to the enzyme cross-react with methyltransferase genes of other modification systems.

Brooks, J.E., *et al.* *Nucleic Acids Research* 19:841-850 (1991), characterizes the cloned *BamHI* restriction modification system from *Bacillus subtilis*. The two genes are divergently oriented and separated by an open reading frame which may serve as a transcriptional regulator in the native bacteria.



- 6 -

Slatko, B.E., *et al.* *Nucleic Acids Research* 15:9781-9796 (1987), describe the cloning, sequencing and expression of the *TaqI* restriction-modification system. These genes have the same transcriptional orientation, with the methylase gene 5' to the endonuclease gene. *E. coli* clones which carry only the endonuclease gene are viable even in the absence of the methylase gene. This is an unusual case possibly explained by the 65°C optimal temperature for *TaqI* restriction and the 37°C optimal temperature for *E. coli* growth.

Howard, K.A., *et al.*, *Nucleic Acids Research* 14:7939-7951 (1986), describe the cloning of the *DdeI* restriction modification system from *Desulfovibrio desulfuricans* by a two step method wherein the methylase gene is first cloned and transformed into *E. coli*, followed by the cloning of the endonuclease gene and transformation of this second gene into the methylase-expressing bacteria. In order to maintain cell viability, high levels of methylase expression are required before the endonuclease gene can be introduced into the bacteria.

Ito, H., *et al.*, *Nucleic Acids Research* 18:3903-3911 (1990), describe the cloning, nucleotide sequence and expression of the *HincII* restriction-modification system. The DNA was isolated from *H. influenzae* Rc, with the two genes positioned in the same transcriptional orientation.

Shields, S.L., *et al.*, *Virology* 76:16-24 (1990), describe the cloning and sequencing of the cytosine methyltransferase gene M.CviJI from the *Chlorella* virus IL-3A. The methylase recognizes the sequence (G/A)GC(T/C/G) and shows amino acid sequence homology with 5-methylcytosine methylases isolated from bacteria. DNA encoding the methylase was obtained from the viral genome which was propagated in the green alga host *Chlorella*.

Xia, Y., *et al.*, *Nucleic Acids Research* 15:6075-6090 (1987), discovered that IL-3A virus infection of *Chlorella*-like green alga induces the expression of the DNA restriction endonuclease CviJI which has novel sequence specificity. This endonuclease recognizes the sequence PuGCPy (wherein Pu =

- 7 -

purine and Py = pyrimidine) but does not cut the sequence PuG<sup>m</sup>CPy, where <sup>m</sup>C is 5-methylcytosine.

U.S. Patent 5,137,823, issued August 11, 1992, to Brooks, J.E., describes a two step method for cloning the *Bam*HI restriction modification system wherein the methylase is cloned first and then introduced into a bacterial host. The endonuclease is then cloned and introduced into the methylase expressing bacteria. This two step procedure provides the host DNA protection from cleavage of the subsequently introduced endonuclease.

U.S. Patent 5,200,333, ('333) issued April 6, 1993, to Wilson, G.G., describes a method for cloning restriction and modification genes. Specifically this reference describes the cloning of the *Taq*I and *Hae*II systems from *Thermus aquaticus* and *Haemophilus aegypticus*, respectively. In this method, bacterial DNA was initially purified and digested, and the fragments were then cloned into a vector to produce a bacterial DNA library. The library was then transformed into *E. coli* and the cells were plated. Colonies were then scraped from the plate to form a primary cell library. Plasmid DNA from this cell library was purified and digested with the endonuclease of the two gene system. Bacteria which expressed the methylase gene had modified plasmid DNA which was protected from endonuclease activity, while plasmids from bacteria which lacked the intact methylase gene were digested. The resulting, undigested plasmid DNA was then transformed into another bacterial strain and the bacteria were plated. Surviving colonies were again harvested to give a secondary cell library and the entire procedure repeated. Plasmids which code for the complete restriction-modification system presumably survived each round of purification and were enriched. Bacteria which survive several rounds of enrichment were subsequently assayed for both methylase and endonuclease activity.

U.S. Patent 5,196,331, ('331) issued March 23, 1993, to Wilson, G.G. and Nwanko, D., describes a method for cloning the *Msp*I restriction and modification genes. This patent describes a method identical to that of U.S.

- 8 -

Patent 5,200,333 ('333). '331 is a continuation-in-part of, and '333 is a continuation of U.S.S.N. 707,079 (now abandoned).

As mentioned above, *Chlorella* virus IL-3A encodes a unique restriction endonuclease called *CviJI* (Xia *et al. Nucleic Acids Res.* 15:6075-6090 (1987)). IL-3A is a large, polyhedral, plaque-forming phycodnavirus (Francki, R.I.B., *et al. Arch. Virol.* suppl.2. Springer-Verlag, Vienna (1991)) that replicates in unicellular, eukaryotic green algae, *Chlorella* strain NC64A (Schuster, A.M., *et al. Virology* 150:170-177 (1986)). The double-stranded DNA genome of IL-3A is approximately 330 kbp (Rohozinski *et al., Virology* 168:363-369 (1989)) and contains 9.7% methylated cytidine (Van Etten, J.L. *et al., Nucleic Acids Res.* 13:3471-3478 (1985)). The cognate methyltransferase of *CviJI*, M.*CviJI*, methylates (A/G)GC(T/C/G) sequences and, has been cloned and sequenced (Shields, S.L. *et al., Virology* 176:16-24 (1990)).

The use of a two/three base recognition endonuclease, such as *CviJI*, to improve numerous conventional molecular biology applications as well as permitting novel applications has been described in co-pending U.S. Patent Application Ser.No. 08/036,481, filed on March 24, 1993. The application discloses methods for generating sequence-specific oligonucleotides from DNA without prior knowledge of the nucleic acid sequence of such DNA, and to methods for cloning and labeling DNA after restriction digestion by a two base recognition endonuclease. The application also teaches methods for generating quasi-random fragments of DNA, methods for cloning, labeling, and sequencing DNA, as well as epitope mapping of proteins. The ability to generate numerous oligonucleotides with perfect sequence specificity or quasi-random distributions of DNA fragments such as is possible with *CviJI*<sup>\*</sup> has important implications for a number of conventional and novel molecular biology procedures.

Infection of *Chlorella* species NC64A with the IL-3A virus produces sufficient *CviJI* restriction endonuclease (*CviJI*) for research purposes. However, production of commercially useful amounts of *CviJI* is limited with this

- 9 -

system due to the slow growth of *Chlorella* algae, the large number of contaminating nucleases associated with the virus, and the small yield of enzyme obtained after purification. In addition, biochemical and biophysical characterization of the enzyme, such as molecular weight determination, are difficult from the native source. Because of these limitations it would be useful to clone the gene for CviJI in order to provide an adequate large scale source of enzyme for use as a molecular biological reagent.

#### SUMMARY OF THE INVENTION

In one of its aspects, the present invention provides purified and isolated polynucleotides (e.g., DNA sequences and RNA transcripts thereof) encoding a unique restriction endonuclease, CviJI, as well as polypeptides and variants thereof which display activities characteristic of CviJI. Activities of CviJI include the recognition of specific DNA sequences, binding to these sequences and cleaving the bound DNA into fragments. Preferred DNA sequences of the invention include viral genomic sequences as well as wholly or partially chemically synthesized DNA sequences. Replicas (i.e., copies of the isolated DNA sequences made *in vivo* or *in vitro*) of DNA sequences of the invention are also contemplated. A preferred DNA sequence is set forth in SEQ ID NO: 2 herein and is contained as an insert in the plasmid pCJH1.4. In another of its aspects, the invention provides purified isolated DNA encoding a CviJI polypeptide by means of degenerate codons.

Also provided are autonomously replicating recombinant constructions such as plasmid DNA vectors incorporating CviJI sequences and especially vectors wherein DNA encoding CviJI or a CviJI variant is operatively linked to an endogenous or exogenous expression control DNA sequence.

According to another aspect of the invention, host cells such as prokaryotic and eukaryotic cells, are stably transformed with DNA sequences of the invention in a manner allowing the desired polypeptides to be expressed

- 10 -

therein. Host cells expressing CviJI and CviJI variant products are useful in methods for the large scale production of CviJI and CviJI variants wherein the cells are grown in a suitable culture medium and the desired polypeptide products are isolated from the host cells or from the medium in which the cells are grown.

5 A preferred host cell is *E. coli*. Still another aspect of the invention is a recombinant CviJI polypeptide.

The present invention is also directed to a method for the digestion of DNA with a restriction endonuclease reagent under conditions wherein said DNA is cleaved at a dinucleotide sequence selected from the group consisting of PyGCPy, PuGCPy, PuGCPu, and wherein Pu = purine and Py = pyrimidine.

10 The present invention is also directed to a method for restriction endonuclease digestion of DNA comprising the step of digesting DNA with a restriction endonuclease reagent under conditions wherein said DNA is digested at 11 of 16 possible dinucleotide sequences and wherein said dinucleotide sequences are selected from the group consisting of PuCGPu, PuCGPy, and PyCGPu, and wherein Pu = purine and Py = pyrimidine.

15 The present invention is directed to shotgun cloning of DNA, epitope mapping, and for labeling DNA using the digestion methods of the present invention. The present invention provides methods for quasi-random fragmenting of DNA using the digestion methods of the present invention under conditions wherein the DNA is only partially cleaved and the site preference of the restriction endonuclease reagent is greatly reduced. By quasi-random is meant an overlapping population of DNA fragments produced by digesting DNA using the methods of the present inventions without apparent site-preference and which appears as a smear upon electrophoresis in a 1-2 wt. % agarose gel. The present invention is also directed to the shotgun cloning and sequencing of quasi-random fragments of DNA produced by the methods of the present invention. Quasi-random fragments in the shotgun cloning method of the present invention are produced by partial digestion of DNA with a restriction endonuclease reagent

- 11 -

according to the methods of the present invention. More particularly, quasi-random fragments of DNA useful in the cloning method of the present invention are produced by the partial digestion of the DNA to be cloned with CviI I, BsuR I or with a restriction endonuclease reagent termed CGase I comprising Taq I and Hpa II. Quasi-random fragments having a length of between about 100 and about 10,000 nucleotides are preferred. More preferred are quasi-random fragments of about 500 to about 10,000 nucleotides in length. The present invention is also directed to the generation of quasi-random fragmentation of DNA using the method of the present invention for the purposes of epitope mapping and gene cloning. These quasi-random fragments are expressed either *in vitro* or *in vivo* and the smallest fragment containing the desired function is identified by screening assays well known in the art.

The present invention is also directed to the production of anonymous primers from any DNA without prior knowledge of the nucleotide sequence. The present invention provides methods for anonymous primer cloning and sequencing after complete digestion of DNA utilizing CviI I, BsuR I or CGase I using the methods of the present invention.

Additionally, the present invention is directed to methods of labeling and detecting DNA comprising the complete digestion of DNA using the methods of the present invention, followed by a heat denaturation step, to yield sequence specific oligonucleotides. In particular, an aspect of the present invention involves labeling DNA with sequence specific oligonucleotides of about 20 to about 200 bases in length (with an average size of between 20-60 bases) generated by CviI I, BsuR I or CGase I digestion of the template DNA.

More particularly, the invention is directed to restriction generated oligonucleotide labeling (RGOL) of DNA which comprises the digestion of an aliquot of template DNA with CviI I followed by a simple heat denaturation step, thereby generating numerous sequence specific oligonucleotides, which can then be utilized for labeling nucleic acids by a number of methods, including primer

- 12 -

extension type reactions with a DNA polymerase and various labels, isotopic or non-isotopic (RGOL-PEL); 5' end labeling with polynucleotide kinase; 3' end labeling using terminal transferase and various labels, isotopic or non-isotopic. Labeling at the 3' end, also referred to as tailing, adds numerous labels per  
5 oligonucleotide (1-200), depending on the labeling conditions. The addition of 10-500 oligonucleotides generated per template, results in a significant signal amplification not obtainable by conventional methods.

The invention is also directed to thermal cycle labeling (TCL) which comprises the simultaneous labeling and amplification of probes utilizing  
10 CviI I or CGase I restriction generated oligonucleotides as the starting material. In this method, natural DNA of unknown sequence is digested with CviI I to generate numerous double-stranded fragments which are then heat denatured to yield oligonucleotides. These oligonucleotides are combined with the intact  
15 template and subjected to repeated cycles of denaturation, annealing, and extension in the presence of a thermostable DNA polymerase or functional fragment thereof which maintains polymerase activity, deoxynucleotide triphosphates and the appropriate buffer. Alpha <sup>32</sup>P-dATP (or any of the other  
three deoxynucleotide triphosphates), biotin-dUTP, fluorescein-dUTP, or digoxigenin-dUTP is incorporated during the extension step for subsequent  
20 detection purposes. Thermal cycle labeling efficiently labels DNA while simultaneously amplifying large amounts of the labeled probe. In addition, TCL probes exhibit a 10 fold improvement in detection sensitivity compared to conventional probes.

The present invention is also directed to TCL in which the  
25 thermostable DNA polymerase supplies endogenous primers for enzymatic extension. This method is referred to as Universal Thermal Cycle Labeling (UTCL). In this method natural DNA of unknown sequence is combined intact with the holo-enzyme of a thermostable DNA polymerase, deoxyribonucleotide triphosphates, and the appropriate buffer. The holo-enzyme and its associated

- 13 -

endogenous primers are then combined with intact template and subjected to repeated cycles of denaturation annealing and extension. Alpha  $^{32}\text{P}$ -dATP,  $^{32}\text{P}$ -dTTP,  $^{32}\text{P}$ -dGTP,  $^{32}\text{P}$ -dCTP, biotin-dUTP, fluorescein-dUTP, or digoxigenin-dUTP is also included in the extension step for subsequent detection purposes.

5 Isotopic labels useful in the practice of the present invention include but are not limited to  $^{32}\text{P}$ ,  $^{33}\text{P}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$  and  $^3\text{H}$ . Non-isotopic labels useful in the present invention include but are not limited to fluorescein biotin, dinitrophenol and digoxigenin.

10 The present invention is also directed to an improved method for purifying CviI I from the algae *Chlorella* infected with the virus IL-3A.

In addition the present invention is directed to restriction endonuclease reagents which, under conditions which relax the sequence specificity of one or more restriction endonucleases, cleave DNA at the dinucleotide sequences AT or TA.

15 The present invention is also directed to a restriction endonuclease reagent comprising in combination, Taq I and Hpa II, which is capable of digesting DNA at 11 of 16 possible dinucleotide sequences, said sequences selected from the group consisting of PuCGPu, PuCGPy, PyCGPy and PyCGPu, and wherein Pu = purine and Py = pyrimidine.

20 The following examples are intended to be illustrative of the several aspects of the present invention and are not intended in any way to limit the scope of any aspect of the present invention.

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

25 Figure 1 is a map of the plasmid p710 which contains DNA sequences encoding for the IL-3A viral methyltransferase M.CviII;

Figure 2 is the nucleotide sequence of 5497 bp of cloned IL-3A viral DNA;



- 14 -

Figure 3 is a restriction map of the cloned IL-3A viral DNA, including the identified open reading frames;

Figure 4 is the DNA sequence of the *Cvi*JI gene with its flanking regions. The predicted amino acid sequence is provided below the nucleotide sequences;

Figure 5A depicts the theoretical frequency and distribution of *Cvi*JI\* restriction generated oligomers of individual lengths; Figure 5B shows the actual frequency and distribution of *Cvi*JI\* restriction generated oligomers of various lengths;

Figure 6 is a flow chart depicting anonymous primer cloning;

Figure 7 is a photographic reproduction of a gel depicting *Cvi*JI restriction digests of pUC19;

Figure 8 is a photographic reproduction of a gel depicting comparisons of sonicated versus *Cvi*JI\* partially digested DNAs;

Figure 9A is a photographic reproduction of an agarose gel electrophoresis analysis of size-fractionated DNA by microcolumn chromatography compared to fractionation by agarose gel electroelution;

Figure 9B-E illustrates additional trials of the same procedures used in Figure 9A;

Figure 10A illustrates the size distribution of DNA fragments produced by partial digestion of DNA by *Cvi*JI and fractionated by microcolumn chromatography;

Figure 10B-C illustrates the size distribution of DNA fragments produced by partial digestion of DNA by *Cvi*JI and fractionated by agarose gel electrophoresis;

Figure 11 is a schematic depiction of the distribution of *Cvi*JI sites in pUC19; and

Figure 12 is a graph of the rate of sequence accumulation by *Cvi*JI\*\* shotgun cloning and sequencing.

- 15 -

**DETAILED DESCRIPTION**

The gene for the restriction endonuclease R.CviJI was cloned into *E. coli* so as to provide an adequate source of R.CviJI for use as a molecular biological reagent. Biologically active CviJI has been purified from *E.coli* to  
5 apparent homogeneity. The molecular weight of *E.coli* derived R.CviJI is 32.5 kD by SDS gel electrophoresis. N-terminal amino acid sequence analysis of this protein and comparison to the nucleotide sequence of the gene revealed that the translation of this enzyme is probably initiated with a GTG start codon, instead of the usual ATG initiation codon. The structural gene is 834 nucleotides in  
10 length coding for a protein of 278 amino acids (31.6 kD). A second peak of R.CviJI activity which elutes separately from the 32.5 kD form can be seen in the initial stages of enzyme purification. Trace amounts of a larger molecular weight form have not been observed to date. However, the R.CviJI gene does possess an in-frame upstream ATG codon which if translated would yield a predicted 41.4  
15 kD protein. The structural gene for this potentially larger product is 1074 nucleotides in length coding for a putative protein of 358 amino acids.

The present invention is also directed to a method for the fragmentation and cloning of DNA using the restriction endonuclease CviJ I under conditions which allow the enzyme to cleave DNA at the dinucleotide sequence  
20 GC. In addition, the present invention is also directed to the cloning of quasi-random fragments of DNA digested using the fragmentation method of the present invention.

As an alternative to the methods for constructing random clone libraries described above, methods were devised for the construction of such  
25 libraries which require fewer steps and reagents, which require smaller amounts of DNA, which have relatively high cloning efficiencies and which takes less time to complete. These methods relate to the recognition that a partial digest with a two or three base recognition endonuclease cleaves DNA frequently enough to be functionally random with respect to the rate at which sequence data may be

- 16 -

accumulated from a shotgun clone bank. The restriction enzyme CviI normally recognizes the sequence PuGCPy and cleaves between the G and C to leave blunt ends (Xia *et al.*, *Nucl. Acids Res.* 15:6075-6090 (1987)). Under "relaxed" conditions (in the presence of 1 mM ATP and 20 mM DTT) the specificity of CviI I can be altered to cleave DNA more frequently and perhaps as frequently as at every GC. This activity is referred to as CviI I\*. Because of the high frequency of the dinucleotide GC in all DNA (16 bp average fragment size for random DNA), quasi-random libraries may be constructed by partial digestion of DNA with CviI I\*. A DNA degradation method with low levels of sequence specificity produces a smear of the target DNA when analyzed by agarose gel electrophoresis. Digestion of the plasmid pUC19 under partial CviI I\* conditions does not result in a non-discrete smear; rather, a number of discrete bands are found superimposed upon a light background of smearing, suggesting that CviI I\* has some site preference. Atypical reaction conditions according to the present invention eliminate this apparent site preference of CviI I\* to produce an activity (termed CviI I\*\*) in combination with a rapid gel filtration size exclusion step, streamlines a number of aspects involved in shotgun cloning.

One aspect of the present invention involves the use of the two/three base recognition endonuclease CviI I, in conjunction with a simple spin-column method to produce libraries equivalent in final form to those generated by the combination of sonication and agarose gel electroelution. However, the method of the present invention requires fewer steps, a shorter time period, and significantly less substrate (nanogram amounts) when compared to conventional procedures. Both small and large sequencing projects using the methods described herein are within the scope of the present invention.

Current sequencing paradigms require the generation of a new template for each 350-500 nucleotides sequenced. On this basis, sequencing both strands of the human genome would require at least 12 million templates 500 nucleotides long, assuming no overlap between templates.

- 17 -

A random approach, such as shotgun sequencing, would require 30 to 50 million templates, assuming the entire genome were randomly subcloned. As many as 250,000 libraries may be needed to generate the requisite templates from a subcloned and ordered array of this genome, depending on the type of vector utilized, and the degree of overlap between such clones. The ability to generate shotgun libraries in a semi-automated, microtiter plate format would greatly simplify such large scale projects.

The development of methods for cloning large DNA molecules in yeast artificial chromosomes (Burke *et al.*, *Science* 236:806-812 (1987), or in bacteriophage P1-derived vectors (Sternberg, *Proc. Natl. Acad. Sci. USA* 87:103-107 (1990)), simplifies the subdivision and analysis of very large genomes. However, the large size of the resulting subclones (100 - 1000 kbp) presents additional challenges for subsequent sequencing efforts. A report of the sequencing of a 134 kbp genome by random shotgun cloning directly into a bacteriophage M13 vector indicates that numerous intermediate stages of subcloning, mapping, and overlapping such clones may be eliminated (Davison, *J. DNA Seq. and Mapping* 1:389-394 (1992). An order of magnitude reduction in the amount of DNA required for shotgun cloning would substantially simplify efforts to directly sequence 100,000 bp sized molecules and beyond.

The ability to generate an overlapping population of randomly fragmented DNA molecules is considered essential for minimizing the closure of nucleotide sequence gaps by the shotgun cloning method. The use of a very frequent-cutting restriction enzyme, such as CviI I, is an approach which has not been utilized. Reaction conditions according to the present invention result in the quasirandom restriction of pUC19 and lambda DNA, as judged by the degree of smearing observed.

The randomness of this CviI I<sup>\*\*</sup> reaction was quantified by sequence analysis of 76 such partially-fragmented pUC19 subclones. The analysis is showed that CviI I<sup>\*\*</sup> partial digestion (limiting enzyme and time) restricts DNA

- 18 -

at PyGCPy, PuGCPu, and PuGCPy (but not PyGCPu), and is thus a hybrid reaction which combines the three base recognition specificity of CviI I with the "two" base recognition specificity of CviI I\*. Interestingly, most of the "relaxed" cleavage observed under CviI I\*\* conditions occurred in those portions of the sequence which were deficient in "normal" restriction sites. CviI I\*\* treatment produces a relatively uniform size distribution of DNA fragments, permitting sequence information to be accumulated in a statistically random fashion.

Shotgun cloning with CviI I\*\* digested DNA is efficient partly because the resulting fragments are blunt ended. Other methods currently used to randomly-fragment DNA, including sonication, DNase I treatment, and low pressure shearing, leave ragged ends which must be converted to blunt ends for efficient vector ligation. Other than a heat denaturation step to inactivate the endonuclease, no additional treatments are required for cloning CviI I\*\* restricted DNA. In addition, the preligation step required to equalize representation of the ends of a DNA molecule prior to sonication or DNase I treatment is not necessary with CviI I\*\* fragmentation. CviI I\* cleaves its cognate recognition site very close to the ends of a linear molecule, as judged by the very small fragments resulting from complete digestion of pUC19 as depicted in Figure 2, lane 1.

The overall efficiency of shotgun cloning depends not only on the fragmentation process, but also upon the size fractionation procedure used to remove small DNA fragments. The efficiency of cloning agarose gel fractionated DNA was found to be unexpectedly variable. Numerous experiments produced an erratic distribution of sized material and the resulting cloned inserts were uniformly small (70% < 500 bp in one trial, 100% < 500 bp in another). The method of the present invention includes a simple and rapid micro-column fractionation method, which has resulted in three to thirteen times more transformants than agarose gel fractionation. More importantly, the size distribution of the cloned inserts from column-fractionated DNA was skewed

- 19 -

toward larger fragments (88% > 500 bp). Micro-column fractionation also eliminates the chemical extraction steps required for agarose fractionated DNA. After the target DNA has been column-fractionated, no further treatments are required for cloning. Combining CviI I<sup>\*\*</sup> partial restriction with micro-column  
5 fractionation permits the construction of useful libraries from as little as 200 ng of substrate, an order of magnitude less starting material than recommended for sonication/end-repair and agarose gel fractionation procedures.

The CviI I<sup>\*\*</sup> reaction represents a unique alternative for controlling the partial digestion of DNA, a technique which is fundamental to the construction  
10 of genomic libraries (Maniatis *et al.* *Cell* 15:687-701 (1978), and restriction site mapping of recombinant clones (Smith, *et al.* *Nucl. Acids Res.* 3:2387-2398 (1976). Partial DNA digests are notably variable and are strongly dependent on the concentration and purity of the DNA, the amount of enzyme used, the incubation time, and the batch of enzyme. Partial digestions may also be variable  
15 with respect to the rate at which a particular recognition sequence is cleaved throughout the substrate. Optimal reaction conditions, such as those which render such partial digests independent of one or more of these variables, allows more precise control of the end product. Several controlling schemes may be employed, including: the addition of a constant amount of carrier DNA (Kohara  
20 *et al.*, *Cell* 50:495-508 (1987)), the use of limiting amounts of Mg<sup>2+</sup> (Albertson *et al.* *Nucl. Acids Res.* 17:808 (1989)), ultraviolet irradiation (Whitaker, *et al.* *Gene* 41:129-134), and the combination of a restriction enzyme and a sequence complementary DNA methylase (Hoheisel *et al.*, *Nucl. Acids Res.* 17:9571-9582 (1989)). Utilizing three different batches of CviI I, and three different DNA  
25 templates from five separate preparations, a uniform CviI I<sup>\*\*</sup> partial digestion pattern was obtained that was primarily time-dependent when a constant ratio of 0.3 units of enzyme per  $\mu$ g of DNA was used.

The rate at which a particular restriction site is cleaved at different locations in a substrate is variable for many endonucleases (Brooks, *et al.*,

- 20 -

*Methods in Enzymol*, 152:113-129 (1987)). Reaction conditions for CviI I may be optimized to substantially reduce the site preferences of this enzyme during partial digestion (see Figure 2, lanes 3 and 4). Normally, "star" reaction conditions result in cleavage at new sites. The use of star reaction conditions according to the present invention (dimethyl sulfoxide [DMSO] and lowered ionic strength) to affect the partial digestion activity of CviI I<sup>\*\*\*</sup> does not result in an altered restriction site cleavage as assayed by sequencing the products of 76 digestion reactions. Instead, the relative rate of cleavage of individual sites appears to be more uniform under these conditions. A 3-5 fold increase in the rate of normal CviI I restriction with the standard buffer and DMSO further substantiates this approach. All of these results indicate that, under the appropriate reaction conditions, CviI I is useful for a number of other applications, such as high resolution restriction mapping and fingerprinting, diagnostic restriction of small PCR fragments, and construction of genomic DNA libraries.

Another aspect of the present invention involves quasi-random fragmentation of DNA using the method of the present invention for epitope mapping and cloning intact genes. The same method as described above for shotgun cloning is utilized, except that an expression vector is used to generate functional proteins from the DNA.

Another aspect of the present invention involves fragmenting DNA using the present invention to generate multiple oligonucleotides from any double-stranded DNA template. Restriction-generated oligonucleotides (RGO) are sequence specific oligonucleotides generated from any DNA according to the present invention. CviI I<sup>\*</sup> presumably cleaves the recognition sequence GC between the G and C to leave blunt ends (Xia *et al.*, *Nucl. Acids Res.* 15:6075-6090, (1987)). Because of the high frequency of dinucleotide GC in all DNA (16bp average fragment size for random DNA), a complete CviI I<sup>\*</sup> restriction results in numerous fragments which are about 20-200 bp in size. These

- 21 -

restriction fragments are generated from an aliquot of the template itself and are heat-denatured to yield numerous single-stranded oligonucleotides which are of variable length but which are specific for the cognate template. Complete CviI I\* restriction of the small plasmid pUC19 (2689 bp) theoretically yields 314 oligonucleotides after a heat-denaturation step. The ability to generate numerous oligonucleotides with perfect sequence specificity is an unusual result of the use of this class of enzyme according to the present invention. Such oligonucleotides are uniquely suited for purposes of labeling DNA, as described below.

One application of CviI I\* restriction-generated oligonucleotides is to directly label them using conventional methods. There are several important advantages in using CviI I\* restriction-generated oligonucleotides. Conventional methods employing synthetic oligonucleotides for detection purposes generally use one oligonucleotide containing one or a few labels. A complete CviI I\* digest generates hundreds of oligonucleotides from a given template, depending on the size of the template, and thus makes hundreds of sites available for labeling, regardless of the labeling scheme utilized. These hundreds of sequence specific restriction-generated oligonucleotides have two important advantages over conventional probes used in nucleic acid detection methods. First, the generation of multiple oligonucleotide probes directed at multiple sites in a given target (theoretically, 314 sites in pUC19) provides enhanced detection sensitivities compared to synthetic oligonucleotides which are directed at 1 or a few sites in a target. The numerous labeled restriction-generated oligonucleotides represent a 10-100 fold amplification of the signal for detection compared to the use of a single oligonucleotide. Second, the short length of the restriction-generated oligonucleotides permits more efficient hybridization. This is important for two reasons. First, hybridization times using restriction-generated oligonucleotides is reduced to 1 hr as opposed to an overnight incubation with conventional probes hundreds of nucleotides in length. This is a very important advantage when using oligonucleotide probes in clinical settings. Second, the penetration of probes into



- 22 -

permeabilized cells is a critical issue for *in situ* hybridization procedures. The smaller the probe, the easier the entry into the cell. Thus, the use of multiple oligonucleotide probes generated by the two base cutters greatly improves the sensitivity of *in situ* hybridization, a technique of considerable importance in research and clinical labs. Finally, when using membrane-based hybridization procedures, only small sections of a target nucleic acid are exposed and available for hybridization. Multiple oligonucleotides derived from a cognate template exhibit better detection sensitivities compared to long probes.

Another application of restriction-generated oligonucleotides for labeling is to employ them as primers in a polymerase extension labeling reaction in conjunction with a repetitive thermal cycling regimen of denaturation, annealing, and extension. Thermal Cycle Labeling (TCL) is a method for efficiently labeling double-stranded DNA while simultaneously amplifying large amounts of the labeled probe. The TCL system employs the two base recognition endonuclease CviI I\* to generate sequence-specific oligonucleotides from the template DNA itself. These oligonucleotides are combined with the intact template and subjected to repeated cycles of denaturation, annealing, and extension by a thermostable DNA polymerase from, for example, *Thermus flavus*. A radioactive- or non-isotopically-labeled deoxynucleotide triphosphate is incorporated during the extension step for subsequent detection purposes. The amplified, labeled probes represent a very heterogeneous mixture of fragments, which appears as a large molecular weight smear when analyzed by agarose gel electrophoresis. Primer-primer amplification, a side product of this reaction (produced by leaving out the intact template in the TCL reaction), may result in enhanced detection sensitivity, perhaps by forming branched structures. Biotin-labeled probes generated by the TCL protocol detect as little as 25 zeptomoles ( $2.5 \times 10^{-20}$  moles) of a target sequence. A 50  $\mu$ l TCL reaction yields as much as 25  $\mu$ g of labeled DNA, enough to probe 25 to 50 Southern blots. After 20 cycles of denaturation and extension, biotin-dUTP-incorporated TCL probes may

- 23 -

be routinely detected at a  $1:10^6$  dilution, which is 1000 fold more sensitive than RPL, and indicates that a significant degree of net synthesis or amplification of the probe is occurring. In addition, non-isotopically-labeled TCL probes exhibit a 10-fold improvement in detection sensitivity when compared to RPL-generated probes.  $^{32}\text{P}$ -labeled probes generated by the TCL protocol may also detect as little as 50 zeptomoles ( $2.5 \times 10^{-20}$  moles) of a target sequence. As little as 10 pg of template DNA is enough to synthesize 5-10 ng of radioactive version of TCL generates probes having extremely high specific activities, e.g. (about  $5 \times 10^9$  cpm/ $\mu\text{g}$  DNA), which permits 5 to 10-fold lower detection limits than conventional labeling protocols.

There are several advantages to using restriction-generated oligonucleotides for primer extension labeling of DNA. One advantage is the specificity of the primers. All of the oligonucleotides generated by the TCL system are specific for the template utilized, unlike random primer labeling (RPL) which utilizes synthetic oligonucleotides 6-9 bases in length having a random sequence. The amount of primer required for efficient labeling with the TCL system is only 10 ng, compared to the 10  $\mu\text{g}$  of random primers utilized for RPL. Due to their short length, random primers anneal very inefficiently above 25-37°C, thus RPL is limited to DNA polymerases such as Klenow or T7. The size of the restriction-generated oligonucleotides are longer than the random primers, which extends the hybridization and extension conditions to include a wide variety of temperatures and polymerases. Thus, the use of the restriction-generated sequence-specific oligonucleotides results in more efficient hybridization and extension as compared to RPL. The TCL system has been optimized for labeling with a thermostable DNA polymerase which allows the option of temperature cycling. After 20 cycles of denaturation and extension, a significant amount of amplified TCL probes can be generated. Most importantly, TCL-labeled probes exhibit a 10 fold improvement in detections sensitivity when compared to RPL-generated probes.

- 24 -

Another aspect of the present invention involves a variation of TCL called Universal Thermal Cycle Labelling (UTCL) in which the extension primers are not supplied by CviJI restriction, but rather, are found endogenously in the enzyme preparations of thermostable DNA polymerases. Random sequence DNA is usually co-purified along with the holo-enzyme preparation of the thermostable DNA polymerases, regardless of the source of the enzyme, i.e. native or cloned. However, only the holo-enzyme, and not the exonuclease minus deletion variants, contain the endogenous DNA. Typically, when the holo-enzymes of thermostable polymerases are used in protocols such as the polymerase chain reaction, the presence of such primers can create spurious results. Methods for circumventing the problems of endogenous DNA are described in *PCR Protocols: A Guide to Methods and Applications*, Eds. M. Innis, *et al.*, Academic Press, 1990.

This residual DNA is rather short (approximately 5-25 bases), as assayed by end-labeling with  $\gamma^{32}\text{P}[\text{ATP}]$  and polynucleotide kinase and acts as endogenous "random" primers in a TCL-type reaction. UTCL combines the holo-enzyme of a thermostable polymerase from, for example, *Thermus flavus*, with the intact DNA template and is subjected to repeated cycles of denaturation, annealing, and extension. A radioactive- or non-isotopically-labeled deoxynucleotide triphosphate is incorporated during the extension step for subsequent detection purposes. The amplified, labeled probe represents a very heterogeneous mixture of fragments, which appears as a large molecular weight smear when analyzed by agarose gel electrophoresis. Biotin-labeled probes generated by the UTCL protocol detect as little as 25 zeptomoles ( $2.5 \times 10^{-20}$  moles) of a target sequence. A 15  $\mu\text{l}$  UTCL reaction yields as much as 5-10  $\mu\text{g}$  of labeled DNA, enough to probe 5 to 10 Southern blots. After 20 cycles of denaturation and extension, biotin-dUTP-incorporated UTCL probes may be routinely detected at a  $1:10^6$  dilution, which is 1000 fold more sensitive than RPL, and indicates that a significant degree of net synthesis or amplification of the probe is occurring. In addition, non-isotopically-labeled UTCL probes exhibit

- 25 -

a 10-fold improvement in detection sensitivity when compared to RPL-generated probes.  $^{32}\text{P}$ -labeled probes generated by the UTCL protocol may also detect as little as 50 zeptomoles ( $2.5 \times 10^{-20}$  moles) of a target sequence. The radioactive version of UTCL generates probes having extremely high specific activities, e.g.  
5 (about  $5 \times 10^9$  cpm/ $\mu\text{g}$  DNA), which permits 5 to 10-fold lower detection limits than conventional labeling protocols.

The present invention is illustrated by the following examples relating to the isolation of a full length viral DNA clone encoding R.CviJI, to the expression of R.CviJI DNA in *E.coli* strain DH5 $\alpha$ F<sup>'</sup>MCR and to purification of  
10 R.CviJI from this bacterial stain. More particularly, Example 1 provides for the propagation of IL-3A virus and isolation of viral genomic DNA. Example 2 addresses the improved expression of a clone for the viral methylase M.CviJI. Example 3 describes the strategy for isolating and cloning the viral R.CviJI gene by a forced co-cloning strategy of the M.CviJI gene. Example 4 describes the  
15 sequencing of cloned IL-3A genomic DNA and identification of the R.CviJI gene. Example 5 relates the methods for purification of CviJI to homogeneity from an *E.coli* strain, DH5 $\alpha$ F<sup>'</sup>MCR, transformed with a plasmid which encodes the R.CviJI enzyme. Example 6 details the amino acid sequence analysis of the purified R.CviJI enzyme. Example 7 describes the analysis of CviJI\* recognition  
20 sequences. Example 8 relates to a technique for producing restriction generated oligonucleotides using CviJI. Example 9 relates the generation of anonymous primers using CviJI. Example 10 describes end-labeling of CviJI restriction generated oligonucleotides. Example 11 describes primer extension labeling of DNA using restriction generated oligonucleotides. Example 12 relates the use of  
25 CviJI in thermal cycle labeling of DNA as well as the method of universal thermal cycle labelling. Example 13 provides a method for generation of quasi-random DNA fragments using CviJI. Example 14 describes fractionation of CviJI digested DNA by size using spin column chromatography. Example 15 details the relative cloning efficiency of CviJI digested, size-fractionated DNA by gel elution and

- 26 -

chromatographic methods. Example 16 describes the comparison of cloning efficiency using lambda DNA fragmented by both sonication and CviJI<sup>\*\*\*</sup> techniques. Example 17 details the use of CviJI<sup>\*\*\*</sup> fragmentation for shotgun cloning and sequencing. Example 18 describes the shotgun cloning of lambda DNA using CviJI. Example 19 describes the use of CviJI in epitope mapping techniques. Example 20 describes the restriction endonuclease reagent CGase I.

### Example 1

#### Propagation of IL-3A Virus

The exsymbiotic *Chlorella-like* alga, NC64A, originally isolated from *Paramecium bursaria* (Karakashian, S.J. and Karakashian, M.W., *Evolution and Symbiosis in the Genus Chlorella and Related Algae. Evolution* 19:368-377 (1965)), was grown and maintained in Bold's basal medium (BBM), (Nichols, H.W. and Bold, H.C. *J. Phycol.* 1:34-38 (1965)) modified by the addition of 0.5% sucrose, 0.1% protease peptone, and 20 µg/ml tetracycline (MBBM). Cultures were inoculated with  $1 \times 10^6$  algae cells/ml and grown at 25°C in 250 ml of MBBM in 500 ml Erlenmeyer flasks on a rotary shaker (150 rpm) in continuous light (ca. 30 µEi, m<sup>-2</sup>,sec<sup>-1</sup>). Growth was monitored by light scattering measured as A<sub>640nm</sub> and/or by direct cell counts with a hemocytometer.

When the cultures reached approximately  $1 \times 10^7$  algae cells/ml they were inoculated with filter sterilized (0.4 µm nitrocellulose filter, Nucleopore, Pleasanton, California) IL-3A virus at a multiplicity of infection of 0.01 and incubated for an additional 48 - 72 hours at 25°C. The crude lysate was then centrifuged at 3000 rpm (2000 xg) for 10 minutes to remove cellular debris. Nonidet P-40 was then added to 1% (v/v) and the virus was pelleted from the supernatant by centrifuging at 15,000 rpm at 4°C for 75 minutes in a Beckman No. 30 rotor. The viral pellet was gently resuspended in 0.05 M Tris-HCl, pH

- 27 -

7.8, and the sample was layered on linear 10 - 40% sucrose gradients equilibrated with 0.05 M Tris-HCl, pH 7.8, and centrifuged for 20 minutes at 20,000 rpm at 4°C in a Beckman SW28 rotor. The viral band, which was present in the center of the gradient as an opaque band, was removed, diluted with 0.05 M Tris-HCl, pH 7.8, and pelleted by centrifugation at 15,000 rpm at 4°C for 120 minutes in a Beckman No. 80 rotor. The virus was resuspended in a small volume (10ml) of 0.05 M Tris-HCl, pH 7.8, and stored at 4°C.

IL-3A viral DNA was purified from the viral particles using a modification of the protocol described by (Miller, S.A., Dykes, D.D., and Polesky, H.I., *Nucleic Acids Res.* 16:1215 (1988)). Briefly, 100 µl of IL-3A virus ( $9.8 \times 10^{11}$  plaque forming units/ml) was diluted with 400 µl of water and then mixed with 10 µl TEN (0.5 M Tris-HCl, pH 9.0, 20 mM EDTA, 10 mM NaCl) and 10 µl of 10% SDS. After incubating at 70°C for 30 minutes the solution was extracted twice with phenol-chloroform-isoamyl alcohol, extracted once with chloroform, and precipitated with ice-cold ethanol using methods well known in the art and resuspended in 500 µl of H<sub>2</sub>O. (Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (Eds.) (1987) *Current Protocols in Molecular Biology*, Wiley, New York; Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989), *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York).

## Example 2

### CviJI Methyltransferase Clone

The CviJI methyltransferase gene (M.CviJI) from *Chlorella* virus IL-3A was cloned and sequenced by Shields *et al.*, *Virology* 176:16-24 (1990). Briefly, *Sau3A* partial digest of *Chlorella* virus IL-3A was ligated to *Bam*HI digested pUC19 and transformed into *E. coli* strain RR1. This library of plasmids was restricted with *Hind*III (AAGCTT) and *Sst*I (GAGCTC), both of which are

- 28 -

inhibited by 5-methylcytidine (5mC) in the AGCT portion of their recognition sequences, and transformed again into RR1 cells. M.CviJI methylates the internal cytidine in (G/A)GC(T/C/G) sequences. If the M.CviJI gene is cloned and expressed appropriately, the plasmid DNA would be expected to be resistant to *HindIII* and *SstI* restriction.

The CviJI methyltransferase gene was originally cloned as a 7.2 kb insert, termed pIL-3A.22. Plasmid pIL-3A.22 was only partially resistant to CviJI digestion. Partial digestion is most likely due to the inefficient expression of the M.CviJI gene and the numerous CviJI sites in both the vector (pUC19 has 45 CviJI sites) and in the insert DNA. The M.CviJI gene was eventually sublocalized to a region of 3.7 kb by subcloning using methods well known in the art (Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (Eds.) (1987) *Current Protocols in Molecular Biology*, Wiley, New York; Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989), *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York ) and testing the subcloned DNA for sensitivity/resistance to *HindIII*, *SstI*, and CviJI. (Shields *et al.*, *supra*) The entire sequence was determined and three open reading frames which could code for polypeptides 161, 367, and 162 amino acids, respectively, were identified. The 367 amino acid open reading frame (ORF) was identified as the M.CviJI gene by three criteria: (i) it is the only ORF located in the region identified by transposon mutagenesis; (ii) it has amino acid motifs similar to those of other cytosine methyltransferases; and (iii) a 1.6 kb *DraI* fragment containing the 367 amino acid ORF (1101 bp) produces the methyltransferase. This 1.6 kb M.CviJI encoding fragment was subcloned into the *EcoRV* site of pBluescript KS(-) (Stratagene, LaJolla, CA), in the same translational orientation as the *lacZ'* gene of this vector. A physical map of the resulting plasmid termed p710 is shown in Figure 1.

- 29 -

The plasmid p710 was digested with several endonucleases to indirectly test the efficiency of *M. Cvi*II expression. Fully active methylase should render the plasmid DNA completely resistant to digestion by the following enzymes: *Hae*III (which recognizes the sequence GGCC), *Sac*I (which recognizes the sequence GAGCTC), and *Hind*III (which recognizes the sequence AAGCTT). The plasmid was partially resistant to *Hae*III (90%) and *Sac*I (90%), and even less resistant to *Hind*III (25%) digestion. This lack of complete protection of the plasmid DNA made it impractical to attempt cloning the three/two base restriction endonuclease encoded by the *R. Cvi*II gene. Thus, improvements in the efficiency of *M. Cvi*II expression were required before attempting to clone the *R. Cvi*II gene.

The translation efficiency of the *M. Cvi*II gene was improved by removing extraneous 5' open reading frames, creating a perfect fusion of the *lacZ'* Shine-Delgarno sequence with the methyltransferase start codon (see Figure 1). This was achieved by site-specific oligonucleotide mutagenesis, using the oligomer

5'-CAATTTACACAGGAAACAGCTATGTCCTTTTCGCACGTTAGAAC-3' (SEQ ID NO: 1) to precisely remove the intervening *lacZ'* DNA. The relevant DNA sequences are indicated in Figure 1 (SEQ ID NO:12). The mutagenesis was facilitated by converting the double stranded plasmid DNA of p710 to single-stranded DNA by co-infecting the *E. coli* host strain with the helper phage R408 (Russel, M., Kidd, S. and Kelly, M.R. *Gene* 45:333-338), using methods well known in the art. The mutagenesis reaction was completed using a commercially available kit according to the manufacturer's instruction (Mutagene, Bio-Rad, Hercules, California). The oligonucleotide was annealed to the single-stranded plasmid, extended in the presence of T4 DNA polymerase, ligated using T4 DNA ligase, and transformed into competent SURE™ cells (Stratagene, La Jolla, California). Transformed cells were then grown overnight as a pool, the DNA isolated and purified.



- 30 -

Enrichment for the mutagenized plasmids was made possible by virtue of the loss of an *Xho*I site located in the sequence that was deleted by mutagenesis. Enrichment was accomplished by digesting the isolated, purified plasmid DNA with *Xho*I, followed by dephosphorylation with calf intestinal alkaline phosphatase (CIAP), and transformed into SURE cells. Plasmid DNA was isolated from 18 individual colonies and the DNA tested for resistance to *Xho*I. Plasmid DNA from 11 colonies were resistant to *Xho*I digestion, indicating that they lacked the deleted sequence. Five of these plasmids were restricted with *Hae*III, *Hind*III, *Pvu*II (which recognizes the sequence CAGCTG), and *Cvi*JI. All five appeared 100% resistant to these enzymes. Four of the plasmids were sequenced and the deletion was confirmed as being correct. One of these, pBMC5, was chosen for further modification.

### Example 3

#### Forced Co-Cloning of R.*Cvi*JI

The location of the R.*Cvi*JI gene on the IL-3A virus genome was inferred as being 3' to the M.*Cvi*JI gene for two reasons: 1) the cloned DNA sequence 5' to the M.*Cvi*JI gene did not produce a restriction activity; and 2) several attempts to clone the DNA 3' to the M.*Cvi*JI gene resulted in deletions/rearrangements of this downstream region. This information permitted a forced co-cloning strategy to obtain the restriction endonuclease gene. This strategy uses a deletion derivative of pBMC5 lacking the 3' half of the M.*Cvi*JI gene. Digestion of the IL-3A genome with the same enzyme used to create the M.*Cvi*JI deletion, followed by ligation of the respective DNAs, transformation, and digestion with enzymes incapable of recognizing methylated DNA (e.g., *Hae*III, *Hind*III, *Pvu*II, *Cvi*JI, etc.) should force the selection of clones which have a restored M.*Cvi*JI gene (and thus active methylase enzyme), as well as downstream DNA. Thus, if a clone is found to be *Cvi*JI resistant, the 3' half of

- 31 -

*M.Cvi*II must have been restored, and downstream DNA containing the *R.Cvi*II gene, at least in part, would presumably be cloned.

The details of this cloning strategy are as follows. pBMC5 has two *Eco*RI sites, one approximately in the middle of the *M.Cvi*II gene, while the other site lies in the vector DNA, 3' to the *M.Cvi*II gene (see Figure 1). pBMC5 was restricted with *Eco*RI and ligated at a dilute concentration (10-50 ng/ $\mu$ l) to favor circularization without the 3' *M.Cvi*II fragment. The reaction mixture was then transformed into competent SURE cells and plated on TY agar containing ampicillin. Plasmid DNA from the resulting colonies was tested for the lack of this *Eco*RI fragment by digestion with *Eco*RI. One of these clones, pBMC5RI, was used for the subsequent co-cloning work. Plasmid pBMC5RI was digested with *Eco*RI and dephosphorylated using CIAP. IL-3A genomic DNA was then digested to completion with *Eco*RI. The *Eco*RI digested pBMC5RI and IL-3A DNAs were combined at a ratio of 1:3 in a ligation reaction using T4 DNA ligase, and the products of the ligation reaction were subsequently used to transform competent SURE cells. The pBMC5RI/IL-3A transformants were not plated, but rather grown overnight in culture as a library or pool of cells. The cells were harvested the next day and DNA was isolated and purified. Isolated, purified DNA was digested with *Hae*III, dephosphorylated with CIAP, and transformed into competent SURE cells. The cells were then plated and grown overnight. Six colonies grew, of which only one containing the plasmid, pCJH1.4, was resistant to *Hae*III. The plasmid pCJH1.4 was found to encode *Cvi*II restriction activity. Plasmid pCJH1.4 was further characterized to localize the gene for *Cvi*II by deletion analysis, subcloning experiments, and sequencing. The plasmid pCJH1.4 was deposited with the American Type Culture Collection on June 30, 1993 under Accession Number 69341.

- 32 -

## Example 4

## Sequencing of Cloned IL-3A DNA Containing CviJI Gene

The *EcoRI* fragment cloned into pCJH1.4 (as described in Example 3) is 4901 bp in length. Except for the 519 bp corresponding to the 3' portion of the *M.CviJI* gene, the remainder of the 4901 bp *EcoRI* fragment cloned into pCJH1.4 was sequenced using the SEQUAL DNA Sequencing System (CHIMERx, Madison, WI) by methods well known in the art. Sequencing was accomplished using three approaches: 1) primer walking on pCJH1.4, 2) cloning various restriction endonuclease digests of pCJH1.4 into an M13 type sequencing vector; and 3) sequencing various restriction endonuclease deletion derivatives of pCJH1.4. The nucleotide sequence of 5497 bp of IL-3A viral DNA is shown in Figure 2 and set forth in SEQ ID NO.: 2.

Six open reading frames (ORF) of 1155 bp (ORF1), 468 bp (ORF2), 555 bp (ORF3), 1086 bp (ORF4), 397 bp (ORF5) and 580 bp (ORF6) which could code for polypeptides containing 358 (41.4 kD), 156 (19.4 kD), 185 (20.3 kD), 362 (38.9 kD), 132 (14.5 kD) and 193 (21.9 kD) amino acids, respectively, were identified (see Figure 3). ORFs 4-6 do not code for the *R.CviJI* gene, as the deletion derivative pCdA12, which lacks the DNA between the *AvaI* and *BamHI* sites (see Figure 3), does produce *CviJI* restriction endonuclease activity. In addition, the deletion derivative pCdEB7, lacking the DNA between the *EcoRI* and *BamHI* sites, did not produce *CviJI* activity. Thus ORF1 or ORF3 were the most likely candidates for encoding the *R.CviJI* gene. The sequence of the 1155 bp ORF1 (SEQ ID NO: 3), its deduced amino acid sequence (SEQ ID NO: 4) (as shown in capital letters), plus flanking bases, is presented in Figure 4. The vertical line in Figure 4 and the associated arrow indicate where the DNA sequence from pCJH1.4 diverges from that of pIL-3A.22-8 (Shields, S.L., *et al.*, *Virology* 76:16-24, 1990). This open reading (ORF1) frame is believed to represent the *CviJI* gene because 14 out of 15 N-

- 33 -

terminal amino acids from the protein sequence (see Example 6) matched the predicted translation product of the nucleic acid sequence (Figure 4). Also, the 32.5 kD molecular weight of the homogeneously purified enzyme described in Example 5 matched the predicted translation product of the nucleic acid sequence (31.6 kD) if the encoded protein was translated beginning at the GTG codon located at nucleotides 299 - 301 (Figure 4), instead of the 5' ATG codon located at nucleotides 59 - 61. This possibility is not surprising in light of the fact that approximately 10% of prokaryotic and eukaryotic gene products begin translation with a GTG start codon, rather than the usual ATG codon (Kozak, M., *Microbiol. Rev.* 47:1-45 (1983); Kozak, M. *J. Cell. Biol.* 108:229 (1989); Gold, L. *et al.*, *Annu. Rev. Microbiol.* 35:365-403 (1981)). The structural gene was identified to be 834 nucleotides in length, coding for a protein of 278 amino acids (31.6 kD) and is set forth in SEQ ID NO: 4. It is also interesting to note that the *CviJI* gene was shown to possess an in-frame, upstream ATG codon which if translated could yield a protein with a predicted molecular weight of 41.4 kD (Figure 4). A larger molecular weight form possessing *CviJI* restriction activity has not been detected by SDS gel electrophoresis. However, a second peak of *CviJI* activity which eluted separately from the 32.5 kD form was detected in the initial stages of enzyme purification. The DNA sequence which could theoretically code for a larger form of *CviJI* would be approximately 1074 nucleotides in length (assuming it starts at the upstream ATG codon) and would code for a protein of 358 amino acids.

#### Example 5

##### Purification of Recombinant *CviJI* Restriction Endonuclease

Initially, 20 ml of LB medium (plus 100 µg/ml ampicillin) were inoculated with a 1 ml stock of *E. coli* transformed with the plasmid pCJH1.4 described above and grown overnight at 37°C with shaking. The next day, 20 ml

- 34 -

of this initial overnight culture was used to inoculate another 1 liter of LB medium and grown overnight. The following day, 50 liters of TB medium (12 g Bacto-Tryptone, 24 g Bacto Yeast Extract, 4 ml glycerol, 2.31 g  $\text{KH}_2\text{PO}_4$ , 12.54 g  $\text{K}_2\text{HPO}_4$ , 0.1 g  $\text{MgSO}_4$ , 100  $\mu\text{g/ml}$  ampicillin, and water to 1 liter) were inoculated with an aliquot of the secondary overnight culture and grown at 37°C with 20 liters/min aeration at 200 RPM, until the  $\text{OD}_{595\text{nm}}$  reached 1.0 unit. Vigorous aeration was essential for CviJI expression and a typical yield contained 70 g of cell paste after centrifugation.

The cell pellet was immediately resuspended in lysis buffer A (30 mM Tris-HCl, pH 7.9 at 4°C, 2 mM EDTA, 10 mM beta-mercaptoethanol, 50  $\mu\text{g/ml}$  phenylmethylsulfonyl fluoride (PMSF), 20  $\mu\text{g/ml}$  benzamidine, 2  $\mu\text{g/ml}$  O-phenantroline, 0.7  $\mu\text{g/ml}$  pepstatin) at a volume of 3 ml of buffer A per 1 g of cells. The cell suspension was then passed through a Manton-Gaulin cell disrupter (Gaulin Corporation, Everett, MA) twice and centrifuged for 1 hr (8000 RPM, Sorvall GS3 Rotor) at 4°C. To the supernatant, solid NaCl was added to a final concentration of 200 mM, and 10% polyethyleneimine (PEI) solution slowly added to a final concentration of 1%. The mixture was stirred for 3 hr, and then centrifuged 30 min, at 4°C, 8000 RPM (Sorvall GS3 Rotor). Solid ammonium sulfate was then added to the supernatant at 0.5 g/ml and the mixture was stirred overnight at 4°C. The precipitated proteins were centrifuged for 1 hr. (8000 RPM, Sorvall GS3 Rotor) at 4°C and the resulting pellet dissolved in 100 ml of buffer B (10 mM  $\text{K}_2\text{PO}_4$ , pH 7.2, 0.5 mM EDTA, 10 mM beta-mercaptoethanol, 50 mM NaCl, 10% glycerol, 0.05% Triton X-100, 50  $\mu\text{g/ml}$  PMSF, 20  $\mu\text{g/ml}$  benzamidine, 2  $\mu\text{g/ml}$  o-phenanthroline, 0.7  $\mu\text{g/ml}$  pepstatin). The dissolved protein solution was then dialysed (14kD cut-off) for 12 hours against three 1 liter changes of buffer B. The dialyzed solution was then diluted to 600 ml with buffer B and applied to a 5 x 20 cm phosphocellulose P11 (Whatman) column (flow rate 100 ml/hr).

- 35 -

The column was then washed with 1.5 liter of buffer B followed by a 0 - 1.5 M NaCl gradient in buffer B (5 liters). R.CvII eluted at approximately 600 mM NaCl. The active fractions were then pooled and concentrated to 50 ml with a 76 mm Amicon YM10 membrane. The resulting solution was then diluted to 300 ml with buffer C (20 mM Tris-acetate, pH 7.4 at 4°C, 2 mM EDTA, 10 mM beta-mercaptoethanol, 50 mM NaCl, 10% glycerol, 0.01% Triton X-100, 50 µg/ml PMFS, 20 µg/ml benzamidine, 2 µg/ml o-phenanthroline, 0.7 µg/ml pepstatin) and applied to 2.5 x 7 cm Heparin-Sephrose column at a flow rate of 25 ml/hr.

After a 400 ml wash with buffer B, R.CvII was eluted with a 1.5 liter gradient of 0 - 1.3 M NaCl in buffer C. CvII eluted at approximately 400 mM NaCl. The most active fractions were pooled and applied to a 2.5 x 7 cm Blue-agarose column equilibrated in buffer D (20 mM Tris-acetate pH 8.0, 1 mM EDTA, 7 mM beta-mercaptoethanol, 30 mM NaCl, 10% glycerol, 0.01% Triton X-100, 50 µg/ml PMFS, 20 µg/ml benzamidine, 2 µg/ml o-phenanthroline, 0.7 µg/ml pepstatin). After a 500 ml wash with buffer D, CvII was eluted with a 0 - 1.5 M NaCl gradient (1.5 l) in buffer D. Active fractions were dialyzed against buffer G (10 mM K/PO<sub>4</sub> pH 7.0 (4°C), 10 mM beta-mercaptoethanol, 50 mM NaCl, 10% glycerol, 0.01% Triton X-100, 50 µg/ml PMFS, 20 µg/ml benzamidine, 2 µg/ml o-phenanthroline, 0.7 µg/ml pepstatin) and loaded (20 ml/h) onto a ceramic HTP column (American International Chemical, Natick MA) (1.5 x 3 cm), equilibrated in buffer F (20 mM Tris-HCl pH 8.0, 0.5 mM EDTA, 3 mM DTT, 50 mM K-acetate, 5 mM Mg acetate, 50% glycerol). After washing with 100 ml of buffer F, a 400 ml gradient 0 - 0.9 M K/PO<sub>4</sub> in buffer F was run. The HTP column was washed with buffer G, containing 3 mg/ml BSA, then with 1 M phosphate buffer and reequilibrated in buffer G. The active fractions were then pooled and concentrated using a TM10 membrane to a final volume of 3 - 4 ml. This concentrate was then applied to a 2.5 x 95 cm Sephadex G-100 column, equilibrated in buffer E (20 mM Tris-HCl

- 36 -

pH 7.5 (4°C), 5 mM Mg-Acetate, 2 mM EDTA, 10 mM beta-mercaptoethanol, 100 mM NaCl, 5% glycerol, 0.01% Triton X-100, 50 µg/ml PMFS, 20 µg/ml benzamidine, 2 µg/ml o-phenanthroline, 0.7 µg/ml pepstatin) at a flow rate of 6 ml/hr, and 3 ml fractions collected. Active fractions were dialyzed against storage buffer F.

The molecular weight of the purified CviJI was determined by comparison to known protein standards on a denaturing 10% SDS polyacrylamide gel and a single band migrating with an apparent molecular weight of 32.5 kilodaltons was seen indicating that by these criteria, CviJI was purified to homogeneity.

#### Example 6

##### N-Terminal Amino Acid Sequence of R.CviJI

To confirm that the restriction endonuclease encoded by the insert in pCJH1.4 was CviJI the sequence of the first 15 N-terminal amino acids of purified CviJI was determined by the Edman degradation method using an Applied Biosystems (Foster City, CA) 477A Liquid Phase Protein Sequencer with an on-line 120A PTH Analyzer. The results of that analysis are shown in Table 1.

- 37 -

Table 1

## N-Terminal Amino Acid Analysis of CviJI

	Amino Acid #	Retention Time (min)	pmol (Raw)	Pmol (-bkgd)	Pmol (+lag)	Pmol Ratio	Amino Acid ID
5	1	9.17	6.11	3.86	5.10	34.53	THR, MET, ARG, OR LYS
	2	10.32	3.92	1.54	1.82	9.96	GLU
	3	10.33	4.28	2.22	2.18	11.96	GLU
	4	27.37	2.23	1.49	1.72	7.64	LYS
	5	27.35	2.37	1.66	1.67	7.39	LYS
10	6	17.95	3.37	2.76	2.81	9.48	ARG
	7	28.10	3.19	1.73	2.08	6.09	LEU
	8	13.58	3.58	2.11	2.49	12.08	ALA
	9	28.10	3.23	1.68	1.58	4.63	LEU
	10	18.17	0.71	0.78	0.36	1.21	ILE
15	11	10.30	1.65	0.78	0.96	5.26	GLU
	12	9.72	8.03	0.41	1.31	3.25	LYS
	13	8.53	1.54	0.53	0.55	2.97	GLN
	14	18.18	2.19	1.74	1.67	5.63	ARG
	15	26.80	3.33	0.43	-	0.89	ILE

20 Abbreviations used: threonine (THR), methionine (MET), arginine (ARG), lysine (LYS), glutamic acid (GLU), leucine (LEU), alanine (ALA), isoleucine (ILE) and glutamine (GLN).

The results of this analysis confirm that the protein encoded by the DNA insert in pCJH1.4 (ORF1) is CviJI.



- 38 -

The following Examples illustrate some of the unique properties of and important uses for CviJI.

**Example 7**

**Analysis of CviJI<sup>\*</sup> Recognition Sequences**

5           The CviJI<sup>\*</sup> recognition sequence (see Xia, *et al.*, *Nuc.Acids Res.*  
15: 6025-6090, 1987) was deduced by cloning and sequencing CviJI<sup>\*</sup> digested  
pUC19 DNA fragments. A complete CviJI<sup>\*</sup> digest of pUC19 was ligated to an  
M13mp18 cloning derivative for nucleotide sequence analysis. The sequence of  
the entire insert was read in order to determine which sites were or were not  
10 utilized. A total of 100 clones were sequenced, resulting in 200 CviJI<sup>\*</sup> restricted  
junctions, the data for which are compiled in Table 2.

Table 2  
Distribution of CwJI\* Sites as Assayed by Cloning and Sequencing

Classification Group	NGCN Recognition Sequence	CwJI* Sites Found in pUC19 (%)	CwJI* Sites Cleared (%)	Sites Not Cleared (%)	Pu/Py Structure
Normal (N)	A C	AGCC 9 (4.4)	23 (11.5)	1 (0.9)	PuPuPyPy
	GC	GGCC 11 (5.4)	24 (12.0)	1 (0.9)	
	G T	GGCT 10 (4.9)	13 (6.5)	0 (0.0)	
		AGCT 15 (7.3)	35 (17.5)	0 (0.0)	
		45 (22.0)	95 (47.5)	2 (1.7)	
Relaxed (R1)	C C	CGCC 11 (5.4)	11 (5.5)	4 (3.5)	PyPuPyPy
	GC	TGCC 12 (5.9)	13 (6.5)	10 (8.6)	
	T T	TGCT 10 (4.9)	10 (5.0)	5 (4.3)	
		CGCT 22 (10.7)	17 (8.5)	7 (6.0)	
		55 (26.0)	51 (25.5)	26 (22.4)	
Relaxed (R2)	A A	AGCA 16 (7.3)	13 (6.5)	5 (4.3)	PuPyPuPu
	GC	GGCA 8 (3.9)	11 (5.5)	3 (2.6)	
	G G	AGCG 11 (5.4)	12 (6.0)	11 (9.5)	
		GGCG 22 (10.7)	18 (9.0)	8 (6.9)	
		57 (27.8)	54 (27.0)	27 (23.3)	
Relaxed (R3)	C A	CGCA 10 (4.9)	0	12 (10.4)	PyPuPyPu
	GC	TGCA 13 (6.3)	0	19 (16.4)	
	T G	CGCG 10 (4.9)	0	27 (23.3)	
		TGCG 15 (7.3)	0	3 (2.6)	
		48 (23.4)	0	61 (51.6)	
	Total 205		200	116	

- 40 -

The dinucleotide GC is found at 205 sites in pUC19. These GC sites (shown in Table 2) can be divided into four classes based on their flanking Pu/Py structure, the normal recognition sequence (N) and three potential classes of relaxed sites (R2 and R3). As seen in Table 2, the fraction of such NGCN sites which belong to each classification is roughly equal (22.0%-27.8%). A total of 200 *Cvi*JI\* restricted junctions were analyzed by sequencing 100 cloned inserts. If *Cvi*JI\* cleaved at all NGCN sites without sequence preferences, it would be expected that the fraction of each classification should be restricted approximately equally. Instead, most of the sites cleaved by this treatment were found to be normal, or PuGCPy sites (47.5%). R1 (PyGCPy) and R2 (PuGCPu) restricted sites were found at nearly the same frequency (25.5% and 27.0%, respectively). Out of 200 *Cvi*JI\* junctions, no R3 (PyGCPu) restricted sites were found. Thus, *Cvi*JI\* cleaves all NGCN sites except for PyGCPu. As *Cvi*JI\* cleaves 12 out of 16 possible NGCN sites, it may be referred to as a 2.25-base recognition endonuclease.

In addition to the restricted sites, those sites which were not cleaved by *Cvi*JI\* conditions were also compiled for analysis, as shown in Table 2. A total of 116 non-cleaved NGCN sites were found in the 100 inserts which were sequenced. PyGCPu sites represented the largest class of non-cleaved sites (52.6%). In only two cases were PuGCPy sites found not to be cleaved. An approximately equal fraction of R1 and R2 sites were not cleaved as were found cleaved (22.4% versus 25.5% for R1 and 23.3% versus 27.0% for R2). Based on the frequency of cleavage, or lack thereof, a hierarchy of restriction under *Cvi*JI\* conditions is evident, where PuGCPy >> PuGCPu = PyGCPy.

- 41 -

## Example 8

**CviJI<sup>\*</sup> Restriction Generated Oligonucleotides**

Due to the high frequency of CviJI or CviJI<sup>\*</sup> restriction, it is possible to generate useful oligonucleotides by digestion and a heat denaturation step as described above. The size and number of the resulting oligonucleotides are important for subsequent applications such as those described above. If for example, an oligonucleotide is to be used with a large genome, it has to be long enough so that the sequence detected has a probability of occurring only once in the genome. This minimum length has been calculated to be 17 nucleotides for the human genome (Thomas, C.A., Jr. *Prog. Nucl. Acid Res. Mol. Biol.*, 5:315 (1966)). Oligonucleotides used for sequencing or PCR amplification are generally 17-24 bases in length. Oligomers of shorter length will often bind at multiple positions, even with small genomes, and thus will generate spurious extension products. Thus, an enzymatic method for generating oligomers should ideally result in polymers greater than 18 bases in length.

The theoretical number of pUC19 CviJI<sup>\*</sup> restriction-generated oligomers is 314 (157 CviJI<sup>\*</sup> restriction fragments x 2 oligomers/fragment), the size distribution of which is shown in panel A of Figure 5. Most of the expected CviJI<sup>\*</sup> restriction-generated oligomers (about 75%) are smaller than 20 bp. This assumes that CviJI is capable of restricting DNA to very small fragments, the shortest of which would be 2 bp. However, in practice, about 93% of the cloned CviJI<sup>\*</sup> fragments were 20-56 bp in size, and 3% of the fragments generated by CviJI<sup>\*</sup> were smaller than 20 bp (panel B of Figure 5). This suggests that CviJI<sup>\*</sup> is not able to bind or restrict those fragments below a certain threshold length. Since the smallest observed fragment is 18 bp, it may be assumed that this length is the minimal size which can be generated from a given larger fragment. Whatever the reason for this phenomenon, CviJI<sup>\*</sup> treatment of DNA produces a

- 42 -

relatively small range of oligomers (mostly 20-60 bases in length), most of which are a perfect size class for molecular biology applications.

### Example 9 Anonymous Primer Cloning

5                   Primers are critical tools in many molecular biology applications such as PCR, sequencing, and as probes. Anonymous primers are useful as sequencing primers for genomic sequencing projects, as probes for mapping chromosomes, or to generate oligonucleotides for PCR amplification.

10                   The Anonymous Primer Cloning (APC) method is a variation of shotgun cloning in that unknown sequences of DNA are being randomly cloned. However, unlike *Cvi*JI shotgun cloning, wherein a partial *Cvi*JI<sup>\*\*</sup> digest of DNA is cloned, anonymous primer cloning utilizes a complete *Cvi*JI<sup>\*</sup> digest to restrict large DNAs into small fragments 20-200 bp in size. These small fragments are cloned into a unique vector designed for excising the anonymous DNA as labeled  
15                   primers. The strategy for this method is illustrated in Figure 6.

                  As illustrated in Figure 6, the APC strategy reduces large DNAs to small fragments, which are cloned and excised for use as primers. Plasmid pFEM has a unique arrangement of the restriction sites for *Mbo*II and *Fok*I, which permits DNA cloned into the *Eco*RV site to be excised without associated vector  
20                   DNA. This is possible because *Fok*I cleaves 9/13 bases to the left of the recognition site shown in pFEM and *Mbo*II cleaves 8/7 bases to the right of the recognition site shown in pFEM, which is well into the cloned anonymous sequence. After *Mbo*II or *Fok*I restriction, a known flanking primer is annealed (primer 1 or 2) and extended using a DNA polymerase and dNTPs. The primer  
25                   is previously end-labeled, or alternatively, one or more radioactive.

- 43 -

After denaturation of the newly synthesized DNA and separation from its cognate template, the labeled anonymous primer is ready for use in sequencing the original template from which it was subcloned. The presence of the pFEM vector sequence fused to the anonymous sequence does not influence the enzymatic extension of this primer from its unique binding site, as the vector DNA is at the 5' end and the unique sequence is located at the 3' end (all polymerases extend 5' to 3'). Both the top and bottom strand primers may be excised from pFEM due to the symmetrical placement of restriction sites and flanking primer binding sites. Thus, two primers may be derived from each cloning event. APC is particularly well suited to the genomic sequencing strategy of Church and Gilbert *Proc Natl. Acad Sci. USA* 81:1991-1995 (1984), although its utility is not limited thereto.

#### Example 10

##### End Labeling of Restriction-Generated Oligonucleotides

As is clear from the foregoing examples, digesting DNA with CviJI\* provides the ability to generate sequence-specific oligonucleotides ranging in size from 20-200 bases in length with an average length of 20-60 bases. Sequence specific oligonucleotides generated by CviJI\* digestion may be labeled directly at the 5'-end or at the 3'-end using techniques well known in that art.

For example, 5'-end labeling may be accomplished by either a forward reaction or an exchange reaction using the enzyme T4 polynucleotide kinase. In the forward reaction,  $^{32}\text{P}$  from  $[\gamma\text{-}^{32}\text{P}]\text{ATP}$  is added to a 5' end of an oligonucleotide which has been dephosphorylated with alkaline phosphatase using standard techniques widely known in the art and described in detail in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual, 2nd Edition*. Cold Spring Harbor Laboratory Press (1989). In an exchange reaction, an excess of ADP (adenosine diphosphate) is used to drive an exchange of a 5'-terminal phosphate

- 44 -

from the sequence specific oligonucleotide to ADP which is followed by the transfer of  $^{32}\text{P}$  from  $\gamma\text{-}^{32}\text{P}\text{-ATP}$  to the 5'-end of the oligonucleotide. This reaction is also catalyzed by T4 polynucleotide kinase and is described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual, 2nd Edition*. Cold Spring Harbor Laboratory Press (1989).

Homopolymeric tailing is another standard labeling technique useful in the labeling of CviII<sup>\*</sup>-generated sequence specific oligonucleotides. This reaction involves the addition of  $^{32}\text{P}$ -labeled nucleotides to the 3'-end of the sequence specific oligonucleotides using a terminal deoxynucleotide transferase. (Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual, 2nd Edition*. Cold Spring Harbor Laboratory Press (1989)).

Commonly used labeling techniques typically employ a single oligonucleotide directed to a single site on the target DNA and containing one or a few labels. Oligonucleotides generated by the method of the present invention are directed to many sites of a target DNA by virtue of the fact that they are generated from a sample of the target sequence. Thus, the hybridization of multiple oligonucleotides (labeled by the methods described above) allows a significantly enhanced sensitivity in the detection of target sequences. In addition, the short length of the labeled oligonucleotides used in the methods of the present invention allows a reduction in hybridization time from overnight (as is used in conventional methods) to 60 mins.

Although labeling sequence specific oligonucleotides with  $^{32}\text{P}$  is described above, labeling with other radionucleotides, and non-radioactive labels is also within the scope of the present invention.

- 45 -

## Example 11

**Primer Extension Labeling of DNA Using  
Restriction-Generated Oligonucleotides (PEL-RGO)**

Another aspect of the present invention includes methods for labeling DNA which include the generation of oligonucleotide primers by complete digestion with *Cvi*JI\*, followed by heat denaturation. PEL-RGO requires three steps: 1) generating the sequence-specific oligonucleotides by *Cvi*JI\* restriction of the template DNA; 2) denaturation of the template and primer; and 3) primer extension in the presence of labeled nucleotide triphosphates. Plasmid DNA may be prepared by methods known in the art such as the alkaline lysis or rapid boiling methods (Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd Edition). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989)). In addition, the vector should be linearized to ensure effective denaturation. A restriction fragment may be labeled after separation on low melting point agarose gels by methods well known in the art.

In PEL-RGO labeling, template DNA to be labeled is divided into two aliquots; one is used to generate the sequence specific oligonucleotide primers and the other aliquot is saved for the primer annealing and extension reaction. A typical reaction mix for generating sequence-specific oligonucleotides is assembled in a microcentrifuge tube and includes: 100 ng DNA; 2  $\mu$ l 5x *Cvi*JI\* buffer; 0.5  $\mu$ l *Cvi*JI (1u/ $\mu$ l); sterile distilled water to 10  $\mu$ l final volume. *Cvi*JI\* 5X restriction buffer includes: 100 mM glycylglycine (Sigma, St. Louis, Missouri, Cat. No. G2265) pH adjusted to 8.5 with KOH, 50 mM magnesium acetate (Amresco, Solon, Ohio, Cat. No. P0013119), 35 mM  $\beta$ -mercaptoethanol (Mallinckrodt, Paris, Kentucky, Cat. No. 60-24-2), 5 mM ATP, 100 mM dithiothreitol (Sigma, St. Louis, Missouri, Cat. No. D9779) and 25% v/v DMSO, (Mallinckrodt Cat. No. 67-68-5). *Cvi*JI is obtained from CHIMERx (Madison, Wisconsin). The reaction mix is incubated at 37°C for 30 min, followed by the



- 46 -

inactivation of *Cvi*II by heating at 65°C for 10 min. The *Cvi*II\*-restricted DNA may be used directly without further purification, or it may be stored at -20°C for several months for subsequent labeling reactions.

5 After heat-inactivating *Cvi*II, 0.2 µg of the digested and undigested DNA are electrophoresed on a 1.5% agarose gel, using a suitable molecular weight marker for comparison. The *Cvi*II restriction fragments appear as a low molecular weight smear in the 20-200 bp range.

10 By way of example, 1-10 ng of linearized pUC19 was labeled under the conditions described below. A template-primer cocktail was prepared by mixing 10 ng of linearized pUC19 DNA template with 20 ng pUC19 sequence-specific oligonucleotides (prepared as described above) and the mixture is brought to a final volume of 17 µl with sterile distilled water. The template-primer mixture is denatured in a boiling water bath for 2 minutes and immediately placed on ice.

15 The following labeling mixture is then added to the template-primer mix: 2.5 µl 10X labeling buffer (500 mM Tris HCl at pH 9.0, 30 mM MgCl<sub>2</sub>, 200 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 20µM dATP, 20µM dTTP, 20µM dGTP, 0.4% NP-40); 5.0 µl [ $\alpha$ -<sup>32</sup>P] dCTP (3000Ci/mmol, 10µCi/µl New England Nuclear, Catalog No. NEG013H); 0.5 µl *Thermus flavus* DNA polymerase (5u/µl) (Molecular Biology Resources, Milwaukee, Wisconsin); up to 25 µl final volume with distilled water. The reaction was incubated at 70°C for 30 min and then stopped by adding 2µl of 0.5M EDTA at pH 8.0 to the reaction mix.

20 The efficiency of the labeling reaction is gauged by the percentage of radioisotope incorporated into labeled DNA. One microliter of the labeling reaction is added to 99 µl of 10mM EDTA in a microcentrifuge tube. This serves as the source of diluted probe for total and trichloroacetic acid (TCA)-precipitable counts. 2 µl of diluted probe is spotted onto the center of a glass fiber filter disc (Whatman number 934-AH). The disc is then allowed to dry and is then placed in a vial containing scintillation cocktail for counting total radioactivity in a liquid

- 47 -

scintillation counter. Another 2  $\mu$ l aliquot from the diluted probe is added to 1 ml of 10% ice cold TCA followed by the addition of 2  $\mu$ l of carrier bovine serum albumin (BSA). This mixture was then placed on ice for 10 minutes. The precipitate is then collected on a glass filter disc (Whatman No. 934-AH) by vacuum filtration. The filter is then washed with 20ml of ice cold 10% TCA, allowed to dry and is placed in a vial containing scintillation cocktail and counted.

Because primer extension oligonucleotide labeling results in net DNA synthesis, the specific activity of labeled DNA is calculated using the following guidelines.

$$\text{Total cpm incorporated} = \text{TCA cpm} \times 50 \times 27$$

Wherein the factor 50 is derived from using 2  $\mu$ l of a 1:100 dilution for TCA precipitation. The number 27 converts this back to the total reaction volume (which is the reaction volume plus 2  $\mu$ l of stop solution).

$$\text{Synthesized DNA (ng of DNA synthesized)} = \text{theoretical yield} \times \text{fraction of radioactivity incorporated.}$$

$$\text{Theoretical yield (ng of DNA)} = \frac{\mu\text{Ci dNTPs added} \times 4 \times 330\text{ng/nmole}}{\text{specific activity dNTP (Ci/nmole)} = \mu\text{Ci/nmole}}$$

$$\text{Fraction of incorporated label} = \text{TCA precipitated cpm} / \text{total cpm.}$$

$$\text{Specific activity (cpm}/\mu\text{g of DNA)} = \frac{\text{total cpm incorporated} \times 1000}{\text{synthesized DNA} + \text{input DNA}}$$

Wherein 1000 is the factor converting nanograms to micrograms.

- 48 -

By way of example, the following represents the calculation of specific activity for an aliquot of pUC19 DNA labeled using this method. Using 50  $\mu\text{Ci}$  of [ $\alpha$ - $^{32}\text{P}$ ]dCTP in a 25  $\mu\text{l}$  reaction, and if the TCA precipitated cpm is 26192 and total cpm is 102047;

$$\begin{aligned} 5 \quad \text{Total cpm incorporated} &= 26192 \times 50 \times 27 = 3.27 \times 10^7 \text{ cpm} \\ \text{Synthesized DNA (ng of DNA synthesized)} &= \\ \text{Theoretical yield} \times \text{fraction of radioactivity incorporated.} \end{aligned}$$

$$\begin{aligned} \text{Theoretical yield} &= \frac{\mu\text{Ci of dNTPs} \times 4 \times 330}{3000 \mu\text{Ci/nmole}} \\ 10 \quad &= \frac{50 \mu\text{Ci} \times 4 \times 330}{3000} \\ &= 22 \text{ ng} \end{aligned}$$

$$\text{Fraction of label incorporated} = \frac{\text{TCA precipitated cpm}}{\text{Total cpm}} = \frac{26192}{102047} = 0.256$$

$$\begin{aligned} 15 \quad \text{Synthesized DNA} &= 22 \times 0.256 \\ &= 5.6 \text{ ng} \end{aligned}$$

$$\text{Specific activity (cpm / } \mu\text{g)} = \frac{\text{Total cpm incorporated}}{\text{Synthesized DNA} + \text{input DNA}} \times 1000$$

$$\text{Input DNA} = 10 \text{ ng}$$

$$\begin{aligned} 20 \quad \text{Specific activity} &= \frac{3.27 \times 10^7 \times 1000}{5.6 + 10} \\ &= 2.09 \times 10^9 \text{ cpm/} \mu\text{g} \end{aligned}$$

Unincorporated radioactive label may be removed using standard methods well known in the art.

- 49 -

Comparisons were made between PEL-RGO vs RPL under similar conditions, and it was observed that a detection limit of 100 fg was seen using PEL-RGO labeled DNA compared to a detection limit of 500 fg with RPL, using a radiolabeled probe.

5

### Example 12

#### Thermal Cycle Labeling and Universal Thermal Cycle Labeling

10

15

20

Thermal Cycle Labeling (TCL) is a method according to the present invention for efficiently labeling double-stranded DNA while simultaneously amplifying large amounts of the labeled probe. TCL of DNA requires two general steps: 1) generation of the sequence-specific oligonucleotides by CviJI\* restriction of the template DNA; and 2) repeated cycles of denaturation, annealing, and extension in the presence of a thermostable DNA polymerase or a functional fragment thereof which maintains polymerase activity. Optimal results are obtained after 20 such cycles, which is best performed in an automated thermal cycling instrument such as a Perkin-Elmer Model 480 thermocycler. In conjunction with such an instrument, about 1.5 hr. is required to complete this protocol. If a thermal cyclor is not available these reactions may be performed using heat blocks. As few as 5 cycles may yield probes with acceptable detection sensitivities. The generation of sequence specific oligonucleotides for use in this method may also be accomplished using the restriction endonuclease reagent CGase I described in Example 20 or the restriction endonuclease Aci I which has as a recognition sequence CCGC.

25

Non-radioactive labeling of DNA using TCL is accomplished by mixing: 10 pg - 100 ng linearized template, 50 ng CviJI\*-digested primers (prepared as described above), 1.5  $\mu$ l 10X labeling buffer, 0.5  $\mu$ l *Thermus flavus* DNA polymerase (5u/ $\mu$ l) (Molecular Biology Resources, Inc., Milwaukee,

- 50 -

Wisconsin), 1  $\mu$ l of 1mM Biotin-11-dUTP (Enzo Diagnostics, New York, New York), 1.5  $\mu$ l each of dATP, dCTP, and dGTP (2 mM), and 1.0  $\mu$ l 2mM dTTP.

Radioactive labeling of DNA using TCL was accomplished by mixing 10 pg - 100 ng of CviJI generated primers, 10 pg-25 ng of linearized  
5 template, 1.5  $\mu$ l of 10X labeling buffer, 5  $\mu$ l of  $^{32}$ P-dCTP (3000 Ci/mmmole, 10  $\mu$ Ci/ $\mu$ l or 40  $\mu$ Ci/ $\mu$ l), 0.5  $\mu$ l of *Thermus flavus* DNA polymerase (5u/ $\mu$ l), and 0.5  $\mu$ l each of dATP, dGTP, and dTTP (1 mM) was added. The reaction mix was brought to a volume of 15  $\mu$ l with deionized H<sub>2</sub>O, overlaid with mineral oil and  
10 cycled through 20 rounds of denaturation, annealing and extension. A typical cycling regimen employed 20 cycles of denaturation at 91°C for 5 sec, annealing at 50°C for 5 sec and extension at 72°C for 30 sec. The reaction is then terminated by adding 1  $\mu$ l of 0.5M EDTA, pH 8.0. The amplified, labeled probe is a very heterogeneous mixture of fragments, which appears as a smear when analyzed by agarose gel electrophoresis.

15 Universal thermal cycle labeling (UTCL) is a method according to the present invention for efficiently labeling double-stranded DNA while simultaneously amplifying large amounts of labeled probe. UTCL is unique in that no sequence information is required regarding the template. The extension  
20 primers are supplied endogenously via the holo-enzyme of the thermostable DNA polymerase and any anonymous DNA template can be labeled by repeated cycles of denaturation, annealing, and extension in the presence of a labeled deoxynucleotide triphosphate. Optimal results are obtained after 20 such cycles, which is best performed in an automated thermal cycling instrument such as a Perkin-Elmer Model 480 thermocycler. In conjunction with such an instrument,  
25 about 1.5 hr are required to complete this protocol. If a thermal cycler is not available these reactions may be performed using heat blocks. As a few as 5 cycles may yield probes with acceptable detection sensitivities.

Non-radioactive labeling of DNA using UTCL is accomplished by mixing: 10 ng linearized template, 1.5  $\mu$ l 10X labeling buffer, 0.5  $\mu$ l *Thermus*

- 51 -

*flavus* DNA polymerase (5u/ $\mu$ l) (Molecular Biology Resources, Inc., Milwaukee, Wisconsin), 1  $\mu$ l of 1mM Biotin-11-dUTP (Enzo Diagnostics, New York, New York), 1.5  $\mu$ l each of dATP, dCTP, and dGTP (2 mM), and 1.0  $\mu$ l 2mM dTTP.

Radioactive labeling of DNA using UTCL was accomplished by mixing: 10 pg-100 ng of linearized template, 1.5  $\mu$ l of 10X labeling buffer, 5  $\mu$ l of  $^{32}$ P-dCTP (3000 Ci/mmol, 10  $\mu$ Ci/ $\mu$ l or 40  $\mu$ Ci/ $\mu$ l), 0.5  $\mu$ l of *Thermus flavus* DNA polymerase (5u/ $\mu$ l), and 0.5  $\mu$ l each of dATP, dGTP, and dTTP (1 mM) was added. The reaction mix was brought to a volume of 15  $\mu$ l with deionized H<sub>2</sub>O, overlaid with mineral oil and cycled through 20 rounds of denaturation, annealing and extension. A typical cycling regimen employed 20 cycles of denaturation at 91°C for 5 sec, annealing at 50°C for 5 sec and extension at 72°C for 30 sec. The reaction is then terminated by adding 1  $\mu$ l of 0.5M EDTA, pH 8.0. The amplified, labeled probe is a very heterogeneous mixture of fragments, which appears as a smear when analyzed by agarose gel electrophoresis.

15                    Estimation of Bio-11 dUTP incorporation:

In order to estimate the level of incorporation of biotin-11-dUTP into DNA, a serial dilution from 1:10 to 1:10<sup>8</sup> of the labeled probe (free of unincorporated biotin-11-dUTP) is made in TE (10mM Tris, 1mM EDTA, pH 8). A microliter of each dilution is placed on a neutral nylon membrane, and the DNA sample is bound to the membrane either by UV cross linking for 3 min or by baking at 80°C for 2 hr.

20                    The unbound sites on the membrane are blocked using a blocking buffer for 15 min at 25°C. Streptavidin-alkaline phosphatase (Gibco-BRL Gaithersburg, Maryland, Cat. No. 9545A) is added to the blocking buffer (0.058 M Na<sub>2</sub>HPO<sub>4</sub>, 0.017 M NaH<sub>2</sub>PO<sub>4</sub>, 0.068 M NaCl, 0.02% sodium azide, 0.5% casein hydrolysate, 0.1% Tween-20) at a 1:5000 dilution and incubated for a 30 min., and the membrane is rinsed 3 times for 10 min. each with wash buffer (1x PBS [0.058 M Na<sub>2</sub>HPO<sub>4</sub>, 0.017 M NaH<sub>2</sub>PO<sub>4</sub>, 0.068 M NaCl], 0.3% Tween,

- 52 -

0.2% sodium azide), rinsed briefly (5 minutes) with AP buffer (100 mM NaCl, 5 mM MgCl<sub>2</sub>, 100 mM Tris-Cl pH 9.5) and then enough AP buffer containing 4.0 µl/ml nitro blue tetrazolium (NBT) (Sigma Cat. No. N6639), (Sigma Cat. No. B6777), and 3.5 µl/ml of 5-bromo-4-chloro-3-indolyl phosphate (BCIP) was added  
5 in order to cover the membrane. The membrane is left in the dark for approximately 30 minutes or until the reaction is complete. The reaction is stopped by rinsing in 1 X PBS.

#### Detection Sensitivities

<sup>32</sup>P-labeled probes generated by the protocol above described  
10 labelling detect as little as 25 zeptomoles ( $2.5 \times 10^{-20}$  moles) of a target sequence. As little as 10 pg of template DNA is enough to synthesize 5-10 ng of radiolabeled probe, which is sufficient for screening 5 Southern blots. The radioactive versions of TCL and UTCL facilitate extremely high specific activities of labeled probe (about  $5 \times 10^9$  cpm/µg DNA), which permits 5-10 fold lower  
15 detection limits than conventional labeling protocols. The synthesis of higher specific activity probes is probably the net result of the sequence-specific oligonucleotide primers and their increased length when compared to the short random primers used in other labeling methods. In addition, the thermal cycling permits probe amplification.

20 Biotin-labeled probes generated by the TCL and UTCL protocols detect as little as 25 zeptomoles ( $2.5 \times 10^{-20}$  moles) of a target sequence. A 15 µl TCL or UTCL reaction yields as much as 5-10 µg of labeled DNA, enough to probe 5 to 10 Southern blots. Biotin-labeled TCL and UTCL probes provide a 10 fold greater detection sensitivity when compared to RPL biotin probes. In  
25 addition, the thermal cycling permits probe amplification.

Non-radioactive, biotinylated probes labeled by the TCL and UTCL methods were shown to have detection limits that are identical to the radioactive probes. These methods have the advantage of eliminating the need to work with

- 53 -

hazardous radioactive materials without sacrificing sensitivity. In addition, results are obtained from non-isotopic probes in 3-4 hours compared to 3-4 days for radiolabeled probes. The ability to substitute non-radioactive probes for radioactive probes may be very useful to clinical laboratories, which do not use radioisotopes but do need greater detection sensitivities. Research laboratories favor the use of non-isotopic systems if detection sensitivity is not an issue. The non-isotopic labeling version of the TCL and UTCL systems represent a major improvement in labeling DNA probes. Non-radioactive probes generated by the methods of the present invention are also useful in the detection of RNA *in situ*. An advantage of this system is that labeling protocols of the present invention yield highly sensitive non-radioactive probes, and the size of the probes are predominantly in the small molecular weight range and can therefore penetrate the tissue easily, unlike RPL. Because non-radioactive probes labeled using the labeling protocols of the present invention have the same detection limits as do radioactive probes similarly labeled, it is within the scope of this invention to use either radioactive or non-radioactive probes for probing, for example, Southern blots, Northern blots, for *in situ* hybridization for the detection of mRNA or DNA in cells or tissue directly, and for colony or plaque lifts.

### Example 13

#### Quasi-Random Fragmentation of DNA

Shotgun cloning and sequencing requires the generation of an overlapping population of DNA fragments. Therefore, conditions were established for the partial digestion of DNA with *Cvi*II to produce an apparently random pattern, or smear, of fragments in the appropriate size range. Conventional methods for obtaining partially restricted DNA include limiting the incubation time or limiting the amount of enzyme used in the digestion. Initially,



- 54 -

agarose gel electrophoresis and ethidium bromide staining of the treated DNA were utilized to assess the randomness and size distribution of the fragments.

CviJI was obtained from CHIMERx (Madison, Wisconsin). Digestion of pUC19 DNA for limited time periods, or with limiting amounts of CviJI under normal or relaxed conditions, did not produce a quasi-random restriction pattern, or smear. Instead, a number of discrete bands were observed, as shown in Figure 7, lane 3 for the CviJI\* partial digestion of pUC19. Complete digests of pUC19 under normal and CviJI\* buffer conditions are shown in lanes 1 and 2 respectively. These results show that, under these relaxed conditions, CviJI has a strong restriction site preference.

To eliminate the apparent restriction site preferences observed under the partial restriction conditions described above, a series of altered reaction conditions were explored. Conditions of high pH, low ionic strength, addition of solvents such as glycerol or dimethylsulfoxide, and/or substitution of  $Mn^{2+}$  for  $Mg^{2+}$  were systematically tested with CviJI endonuclease using the plasmid pUC19. Figure 7 shows the results of these tests. In Lane M, a 100 bp DNA ladder was run. In Lanes 1-4, pUC19 DNA (1.0  $\mu$ g) was run after digestion at 37°C in a 20  $\mu$ l volume for the following times and conditions: Lane 1, complete CviJI digest (1 unit of enzyme for 90 min in 50 mM Tris-HCl, pH 8.0, 10 mM  $MgCl_2$ , 50 mM NaCl); Lane 2, complete CviJI\* digest (1 unit of enzyme for 90 min in 50 mM Tris-HCl, pH 8.0, 10 mM  $MgCl_2$ , 50 mM NaCl, 1 mM ATP, 20 mM DTT); Lane 3, partial CviJI\* digest (0.25 units of enzyme for 30 min in 50 mM Tris-HCl, pH 8.0, 10 mM  $MgCl_2$ , 50 mM NaCl, 1 mM ATP, 20 mM DTT); Lane 4, partial CviJI\*\* digest (0.5 units of enzyme for 60 min in 10 mM Tris-HCl, pH 8.0, 10 mM  $MgCl_2$ , 10 mM NaCl, 1 mM ATP, 20 mM DTT, 20% v/v DMSO); and Lane 5, uncut pUC19 (1.0  $\mu$ g).

The digestion condition which yielded the best "smearing" pattern was obtained when the ionic strength of the relaxed reaction buffer was lowered and an organic solvent was added (Figure 7, lane 4). Plasmid pUC19 partially

- 55 -

digested under these conditions yields a relatively non-discrete smear. This activity is referred to as *CviJI*<sup>\*\*</sup> to differentiate it from the originally-characterized star activity described in Xia *et al.*, *Nucl. Acids Res.* 15:6075-6090 (1987). The appearance of diffuse, faint bands overlying a background smear  
5 generated from this 2686 bp molecule indicates that some weakly preferred or resistant restriction sites may bias the results of subsequent cloning experiments.

DNA was mechanically sheared by sonication utilizing a Heat Systems Ultrasonics (Farmingdale, New York) W-375 cup horn sonicator as specified by Bankier *et al.*, *Methods in Enzymology* 155:51-93 (1987). DNA  
10 fragmented by this method has random single-stranded overhanging ends (ragged ends).

*CviJI*<sup>\*</sup> digested, and sonicated samples were size fractionated by agarose gel electrophoresis and electroelution, or by spin columns packed with the size exclusion gel matrix, Sephacryl S-500 (Pharmacia LKB, Piscataway N.J.) to  
15 eliminate small DNA fragments. Spin columns (0.4 cm in diameter) were packed to a height of 1.3 cm by adding 1 ml of Sephacryl S-500 slurry and centrifuging at 2000 RPM for 5 minutes in a Beckman CPR centrifuge. The columns were rinsed 3 times with 1 ml aliquots of 100 mM Tris-HCl (pH 8.0) by centrifugation at 2000 RPM for 2 min. Typically, 0.2-2.0  $\mu$ g of fragmented DNA in a total  
20 volume of 30  $\mu$ l was applied to the column. The void volume, containing those DNA fragments larger than 500 bp, was recovered in the column eluant after spinning at 2000 RPM for 5 minutes. The capacity of this micro-column procedure is 2  $\mu$ g of DNA. Agarose gel electrophoresis and electroelution are described in detail by Sambrook *et al.* *Molecular Cloning: A Laboratory Manual*,  
25 Second Edition Cold Spring Harbor Laboratory Press, Cold Spring Harbor N. Y. (1989) and is well known to those skilled in the art. In these experiments, 5  $\mu$ g of sample was pipetted into a 2 cm-wide slot on a 1% agarose gel. Electrophoresis was halted after the bromophenol blue tracking dye had migrated

- 56 -

6 cm. Fragments larger than 750 bp, as judged by molecular size markers, were separated from smaller sizes and electrophoresed onto dialysis tubing (1000 MW cutoff). The fractionated material was extracted with phenol-chloroform and precipitated using ice cold ethanol (50% final volume) and ammonium acetate (2.5 M final concentration).

The ragged ends of the sonicated DNA were rendered blunt utilizing two different end repair reactions. In one end repair reaction (ER 1) sonicated DNA was treated according to the procedure outlined by Bankier *et al. Methods in Enzymology* 155:51-93 (1987), where 2.0  $\mu$ g of sonicated lambda DNA is combined with 10 units of the Klenow fragment of DNA polymerase I, 10 units T4 DNA polymerase, 0.1 mM dNTPs, (deoxynucleotide triphosphates=deoxyadenosine triphosphate, deoxythymidine triphosphate, deoxycytosine triphosphate, and deoxyguanosine triphosphate) and reaction buffer (50 mM Tris-HCl, pH 7.5, 10 mM  $MgCl_2$ , 10 mM DTT). This mixture was incubated at room temperature for 30 min followed by heat denaturation of the enzymes at 65°C for 15 minutes. In a second end repair reaction (ER 2), an excess of the reagents and enzymes described above were utilized to ensure a more efficient conversion to blunt ends. In this reaction, 0.2  $\mu$ g of the sonicated lambda DNA sample was treated under the same reaction conditions described above.

Figure 8 shows comparisons of the size distributions of sonicated DNA versus DNA that was partially digested with *Cvi*II<sup>\*\*\*</sup>. In Lanes M, a 1 kb DNA ladder was run. In Lanes 1-3, untreated  $\lambda$  DNA (0.25  $\mu$ g), sonicated  $\lambda$  DNA (1.0  $\mu$ g), and *Cvi*II<sup>\*\*\*</sup> partially-digested  $\lambda$  DNA (1.0  $\mu$ g) were run, respectively. In Lanes 4-6, untreated pUC19 (0.25  $\mu$ g), sonicated pUC19 (1.0  $\mu$ g), and *Cvi*II<sup>\*\*\*</sup> partially-digested pUC19 (1.0  $\mu$ g) were run, respectively.

Fragmentation of a large substrate such as lambda DNA (45 kb) revealed essentially no banding differences between the *Cvi*II<sup>\*\*\*</sup> method and sonication, as demonstrated in Figure 8, lanes 2 and 3. In addition, pUC19 DNA

- 57 -

that was partially digested with *Cvi*JI<sup>\*\*</sup> gave a size distribution or "smear" that closely resembled that achieved with sonication (Figure 8, lanes 5 and 6). As expected, the minor bias evident with a small molecule such as pUC19 was not detectable with a larger substrate such as lambda DNA.

5                   The intensity and duration of sonic treatment affects the size distribution of the resulting DNA fragments. The results obtained from the sonication of lambda and pUC19 samples (Figure 8) were obtained from three 20 second pulses at a power setting of 60 watts. Sonication-generated smears are similar, although the size distribution of fragments is consistently greater with  
10   *Cvi*JI<sup>\*\*</sup> fragmentation. This result favors the cloning of larger inserts, which facilitates the efficiency of end-closure strategies (Edwards *et al.*, *Genome* 6:593-608 (1990)). The size distribution of the DNA fragmented by *Cvi*JI<sup>\*\*</sup> is controlled by incubation time and amount of enzyme, variables which are readily optimized by routine analysis. An excess of enzyme or a long incubation time  
15   will completely digest pUC19 DNA, resulting in fragments which range in size from approximately 20 bp to approximately 150 bp (Figure 7, lanes 1 and 2). The results shown in Figure 8 were obtained by incubating pUC19 for 40 minutes and lambda DNA for 60 minutes with 0.33 units of *Cvi*JI/ $\mu$ g substrate. The efficiencies of the two methods for randomly fragmenting DNA were  
20   quantitatively analyzed for use in molecular cloning, as described below.

#### Example 14

##### Rapid DNA Size Fractionation Utilizing Spin Column Chromatography

25                   The amount of data obtained by the shotgun sequencing approach is substantially increased if fragments of less than 500 bp are eliminated prior to the cloning step. Small fragments yield only a portion of the sequence data which may be collected from polyacrylamide gel based separations and, thus, such small fragments lower the efficiency of this strategy. Agarose gel electrophoresis

- 58 -

followed by electroelution is commonly used to size fractionate DNA prior to shotgun cloning (Bankier *et al.*, *Methods in Enzymol.* 155:51-93 (1987)). Approximately three hours are required to prepare the agarose gel, electrophorese the sample, electroelute fragments larger than 500 bp, perform phenol-chloroform  
5 extractions, and precipitate the resulting material.

The results of 5 out of 9 independent trials size-fractionating CviJI<sup>\*\*</sup>-fragmented lambda DNA by agarose gel electrophoresis are shown in Figures 9A-E. Figures 9A-D illustrate the following. In Figure 9A: Lane M, 1 kb DNA ladder; lane  $\lambda$ , untreated  $\lambda$  DNA (0.25  $\mu$ g); lane 1, unfractionated  
10 (UF) CviJI<sup>\*\*</sup> partially-digested  $\lambda$  DNA (1.0  $\mu$ g); lane 2, column-fractionated (CF) CviJI<sup>\*\*</sup> partially-digested  $\lambda$  DNA (1.0  $\mu$ g); lane 3, gel-fractionated (GF) CviJI<sup>\*\*</sup> partially-digested  $\lambda$  DNA (1.0  $\mu$ g); and in Figures 9B-E are additional trials of the same treatments as in the lanes of Figure 9A which have the same label.

Small DNA fragments may also be removed by passing the sample  
15 through a short column of Sephacryl S-500. Approximately 15 min. are needed to prepare the column and 5 min. to fractionate the DNA by this method.

The results of three out of nine trials using a Sephacryl S-500 column are shown in Figures 9A-C. The efficiency of eliminating small DNA fragments (<500 bp) by spin column chromatography appears high, and the  
20 reproducibility was excellent. This result is in contrast to the agarose gel electrophoresis and electroelution data presented in Figures 9A-E wherein nine replicate trials of this method yielded nine differently sized products, regardless of the source of the agarose. Both methods yielded 30-40% recoveries as measured by UV spectrophotometry. To quantitate the relative efficiencies of the  
25 two fractionation methods, the lambda DNA size fractionated in Figure 9A lanes 2 and 3, and Figure 9B lane 3 were analyzed for cloning efficiency and insert size, as described below.

- 59 -

**Example 15**  
**Cloning Efficiencies of Gel Elution and**  
**Chromatography Fractionation Methods**

The efficacy of size selection was quantified by two criteria: 1) by  
5 comparing the relative cloning efficiency of *Cvi*II\*\* partially-digested lambda  
DNA fragments fractionated either by agarose gel electrophoresis and  
electroelution or micro-column chromatography, and 2) determining the size  
distribution of the resulting cloned inserts. To reduce potential variables, large  
quantities of the cloning vector and ligation cocktail were prepared, ligation  
10 reactions and transformation of competent *E. coli* were performed on the same  
day, numerous redundant controls were performed, and all cloning experiments  
were repeated twice. Ligation reactions were carried out overnight at 12°C in 20  
μl mixtures using the following conditions: 25 mM Tris-HCl (pH 7.8), 10 mM  
MgCl<sub>2</sub>, 1 mM DTT, 1 mM ATP, DNA, and 2000 units of T4 DNA ligase. For  
15 unfractionated samples, 10 ng of fragments and 100 ng of *Hinc*II-restricted,  
dephosphorylated pUC19 were combined under the above conditions. For  
Sephacryl S-500 fractionated samples, 50 ng of size-selected fragments were  
ligated with 100 ng of *Hinc*II-restricted, dephosphorylated pUC19. This increase  
in fractionated DNA was determined empirically to compensate for the lower  
20 concentration of "ends" resulting from the fractionation procedure and/or the  
lowered efficiency of cloning larger fragments. Ligation reaction products were  
added to competent *E. coli* DH5αF' (φ80*dlacZ*Δ*M15* Δ(*lacZYA-argF*)U169 *deoR*  
*gyrA96 recA1 relA1 endA1 thi-1 hsdR17*(r<sub>K</sub><sup>-</sup>,m<sub>K</sub><sup>+</sup>) *supE44* λ-) in a  
transformation mixture as specified by the manufacturer (Life Technologies,  
25 Bethesda, Maryland) and aliquots of the transformation mixture were plated on  
T agar (Messing, *Methods in Enzymol.* 101:20-78 (1983)) containing 20 μg/ml  
ampicillin, 25 μl of a 2% solution of isopropylthiogalactoside (IPTG) and 25 μl  
of a 2% solution of 5-dibromo-4-chloro-3-indolylgalactoside (X-GAL). The

- 60 -

cloning efficiencies reported are the average of triplicate platings of each ligation reaction. The concentration of the fractionated material was checked spectrophotometrically so that 50 ng was added to all ligation reactions. This material was ligated to *HincII*-digested and dephosphorylated pUC19. This  
5 cloning vector was chosen because it permits a simple blue to white visual assay to indicate whether a DNA fragment was cloned (white) or not (blue) (Messing, *Methods in Enzymol.* 101:20-78 (1983)).

A summary of the cloning efficiencies calculated from two independent trials is given in Table 3.

- 61 -

TABLE 3

**Cloning Efficiencies of CviJI<sup>\*\*</sup> Partially Digested Lambda DNA  
Fractionated by Microcolumn Chromatography Versus Agarose Gel  
Electroelution.**

5		Trial I		Trial II	
		<u>Colony Phenotype</u>			
	<u>DNA/treatment</u>	<u>Blue</u>	<u>White</u>	<u>Blue</u>	<u>White</u>
	Supercoiled pUC19	55000	< 10	50000	< 10
	pUC19/HincII/CIAP	210	< 1	320	1
10	pUC19/HincII/CIAP/ T4 DNA ligase	150	4	210	7
	λ/CviII** partial/CF + pUC19	140	240	210	240
	λ/CviII** partial/GFE1 + pUC19	98	49	200	18
15	λ/CviII** partial/GFE2 + pUC19	82	54	95	74

Cloning efficiencies reflect the number of ampicillin-resistant colonies/ng pUC19 DNA. CIAP represents treatment with calf intestinal alkaline phosphatase used to dephosphorylate *HincII*-digested pUC19 to minimize self-ligation. CF refers to DNA that was fractionated on Sephacryl S-500 columns as described above. GFE1 and GFE2 refer to two runs wherein DNA was fractionated by agarose gel electrophoresis and electroeluted. λ refers to bacteriophage λ DNA.

These trials represent repeated experiments in which λ DNA fragments generated by CviJI<sup>\*\*</sup> partial digestion were ligated to *HincII*-linearized,



- 62 -

dephosphorylated pUC19 and transformed into DH5 $\alpha$  F' competent cells described above. The first three rows in Table 2 show controls performed to establish a baseline to better evaluate the various treatments. Supercoiled pUC19 transforms *E. coli* 10 times more efficiently than the *Hinc*II-digested plasmid and 150-260 times more efficiently than the *Hinc*II-digested and dephosphorylated plasmid. The number of blue and white colonies which resulted from transforming *Hinc*II-cut and dephosphorylated pUC19 was determined both before and after treatment with T4 DNA ligase in order to differentiate these background events from cloning inserts. The background of blue colonies (which represent the uncut and/or non-dephosphorylated population of molecules) averaged 0.4 %, compared to supercoiled plasmid. The background of white colonies (which presumably results from contaminating nucleases in the enzyme treatments or genomic DNA in the plasmid preparations) after *Hinc*II-digestion, dephosphorylation, and ligation of pUC19 averaged 0.014% as compared to the supercoiled plasmid.

The number of white colonies obtained when micro-column fractionated DNA was cloned into pUC19 was 240/ng vector in both trials. The efficiency of cloning gel fractionated and electroeluted DNA ranged from 18-74 white colonies/ng vector. The data show that column fractionated DNA results in three to thirteen times the number of white colonies, and presumably recombinant inserts, as gel fractionated and electroeluted DNA. The size distribution of the inserts present in these white colonies is depicted in Figures 10A-C. In Figure 10A, a *Cvi*II<sup>\*\*\*</sup> partial digest of 2 $\mu$ g of  $\lambda$  DNA was size fractionated on a 4 mm by 13 mm column of Sephacryl S-500 at 2,000 x g for 5 minutes. The void volume containing partially digested DNA was directly ligated to linear, dephosphorylated pUC19 and 43 resulting clones were analyzed for insert size. The DNA for this experiment is the same as that shown in Figure 9A, lane 2. In Figure 10B, a *Cvi*II<sup>\*\*\*</sup> partial digest of 5  $\mu$ g of  $\lambda$  DNA was size fractionated by agarose gel electroelution. The eluted DNA was phenol-extracted and ligated to linear, dephosphorylated pUC19, and the resulting 40 clones were

- 63 -

analyzed for insert size. The DNA for this experiment is the same as that shown in Figure 9A, lane 3. In Figure 10C, the procedure is the same as in Figure 9B, except the DNA for this experiment came from Figure 9B, lane 3.

5 A total of 43 random clones obtained from micro-column chromatography fractionation were analyzed for insert size (as shown in Figure 10A). Most of these inserts were larger than 500 bp (37/43 or 86%), 11.6% (5/43) were smaller than 500 bp, and one clone (2.3%) was smaller than 250 bp. The average insert size was 1630 bp. These results are in contrast to those obtained by agarose gel fractionation (as shown in Figures 10B and 10C). In the  
10 first trial (Figure 10B) most of the inserts were smaller than 500 bp (26/37 or 70.3%) and only 29.7% (11/37) were larger than 500 bp in size. In the second trial (Figure 10C) all of the inserts (40 total) were smaller than 500 bp. Thus, the use of agarose gel electroelution for the size fractionation of DNA results in unexpectedly variable and low cloning efficiencies.

15

#### Example 16

##### Cloning Sonicated and *Cvi*JI<sup>\*\*</sup>-Digested Lambda DNA

To compare the cloning efficiencies of sonicated and *Cvi*JI<sup>\*\*</sup>-digested nucleic acid,  $\lambda$  DNA was fragmented by each of these methods and ligated to pUC19 which was linearized with *Hinc*II and dephosphorylated to  
20 minimize self-ligation.

DNA fragmented by *Cvi*JI<sup>\*\*</sup> digestion and sonication was cloned both before and after Sephacryl S-500 size fractionation. Sonicated lambda DNA was subjected to an end repair treatment prior to ligation. Ligations were performed as described in Example 11. One-tenth of the ligation reaction (2  $\mu$ l)  
25 was utilized in the transformation procedure, and the fraction of nonrecombinant (blue) versus recombinant (white) colonies was used to calculate the efficiency of this process.

- 64 -

The efficacy of the methods was quantified by comparing the cloning efficiency of lambda DNA fragments generated either by sonication or *Cvi*II\*\* partial digestion. To reduce potential cloning differences based on size preference, the size distribution of the DNA generated by these two methods was closely matched. Other experimental details were designed to reduce potential variables, as described above. Certain variables were unavoidable, however. For example, the sonicated DNA fragments required an enzymatic step to repair the ragged ends as described in Example 1 prior to ligation, whereas the *Cvi*II\*\* digests were heat-denatured and directly ligated to *Hinc*II digested pUC19.

A summary of the cloning efficiencies calculated from two independent trials is given in Table 4, section A (unfractionated samples), and Section B (fractionated samples).

**TABLE 4**  
**Cloning Efficiencies of *CvJ1*<sup>\*\*</sup> Partially Digested  $\lambda$  DNA**  
**Versus Sonicated  $\lambda$  DNA**

**A. Unfractionated Samples**

DNA/treatment	Trial I		Trial II	
	Colony Phenotype		Colony Phenotype	
	Blue	White	Blue	White
Supercoiled pUC19	30000	<10	16000	<10
pUC19/ <i>HincII</i> /CIAP	150	<1	31	1
pUC19/ <i>HincII</i> /CIAP/ T4 DNA ligase	100	<1	15	1
$\lambda$ /AluI + pUC19	200	400	73	250
$\lambda$ / <i>CvJ1</i> <sup>**</sup> Partial + pUC19	100	160	97	340
$\lambda$ /Sonicated + pUC19	-	-	11	29
$\lambda$ /Sonicated/ER 1 + pUC19	17	10	10	44
$\lambda$ /Sonicated/ER 2 + pUC19	-	-	40	100

6 6

TABLE 4 (cont'd)

## B. Fractionated Samples

DNA/treatment	Trial I		Trial II	
	Colony Phenotype		Colony Phenotype	
	Blue	White	Blue	White
Supercoiled pUC19	35000	<10	12000	<10
pUC19/ <i>HincII</i> /CIAP	30	<1	180	<1
pUC19/ <i>HincII</i> /CIAP/ T4 DNA ligase	60	<1	10	<1
$\lambda$ /A <sub>1</sub> ul + pUC19	28	23	33	48
$\lambda$ /Cv <sub>1</sub> II <sup>**</sup> Partial + pUC19	31	90	36	68
$\lambda$ /Sonicated + pUC19	20	6	99	19
$\lambda$ /Sonicated/ER 1 + pUC19	27	32	40	19
$\lambda$ /Sonicated/ER 2 + pUC19	-	-	25	63

- 67 -

Cloning efficiencies represent the number of ampicillin-resistant colonies/ng pUC19 DNA. CIAP indicates treatment with calf intestinal alkaline phosphatase. ER 1 and ER 2 are end repair methods described in Example 13.  $\lambda$  refers to bacteriophage lambda.

5           The indicated trials represent repeated experiments in which two identical sets of lambda DNA fragments generated by *AluI* complete digestion, *CviJI*<sup>\*\*</sup> partial digestion, or sonication were each ligated to *HincII*-linearized, dephosphorylated pUC19 and transformed into DH5 $\alpha$ F' competent cells. The cloning efficiencies reported are the average of triplicate platings of each ligation  
10 reaction. In case the Sephacryl S-500 size fractionation step introduced inhibitors of ligation or transformation or resulted in differences attributable to the size of the material, the sonicated and *CviJI*<sup>\*\*</sup>-digested samples were ligated with pUC19 both prior to (A) and after (B) the fractionation steps. The first three rows in Table 4, sections A and B, are controls performed to establish a baseline to better  
15 evaluate the various treatments. These data show that supercoiled pUC19 transforms *E. coli* 200-1000 times more efficiently than the *HincII*-restricted and dephosphorylated plasmid. Without this dephosphorylation step, the cloning efficiency is 10% that of the supercoiled molecule (data not presented). The background of blue colonies averaged 0.5% in these experiments, compared to  
20 supercoiled plasmid, while the background of white colonies averaged 0.005%.

A comparison of the data from unfractionated versus fractionated samples in Table 4, sections A and B, reveals a general decline in the number of white and blue colonies obtained after sizing. This decrease is primarily due to the fact that cloning efficiencies are dependent upon the size of the fragment,  
25 favoring smaller fragments and thus giving higher efficiencies for the unfractionated material. This is illustrated by comparing the efficiency of cloning unfractionated and fractionated  $\lambda$  DNA which was completely restricted with *AluI*. This four base recognition endonuclease produces blunt ends and cuts  $\lambda$  DNA (48,502 bp) at 143 sites. Only 25 of the resulting 144 fragments (17%) are larger

- 68 -

than 500 bp. The number of white colonies obtained when unfractionated  $\lambda$  DNA, completely restricted with *AluI*, was cloned into pUC19 ranged from 250-400/ng vector, versus 23-48/ng vector for the fractionated material. This ten fold decrease was only noticed for the  $\lambda$  *AluI* digests, and probably reflects the large  
5 portion of small molecular weight fragments (approximately 75%) which is excluded from the fractionated ligation reactions.

The number of white colonies obtained when unfractionated *CviJI*<sup>\*\*</sup> treated  $\lambda$  DNA was cloned into pUC19 ranged from 160-340/ng vector, versus 68-90 white colonies/ng vector if the same material was fractionated. Unfractionated  
10  $\lambda$  DNA, completely digested with *AluI*, results in cloning efficiencies very similar to unfractionated *CviJI*<sup>\*\*</sup> treated DNA. Sonicated  $\lambda$  DNA is a poor substrate for ligation, compared to *CviJI*<sup>\*\*</sup> treatment, as indicated by the roughly ten-fold reduced cloning efficiencies.

Enzymatic repair of the ragged ends produced by sonication results  
15 in an increased cloning efficiency. Using conditions described in Example 13 for the first end repair treatment (ER 1), 10-44 (fractionated) and 19-32 (unfractionated) white colonies/ng vector were observed. However, ER 1 conditions may not be optimal, as an alternate end repair reaction (ER 2) (as described in Example 13) resulted in greater numbers of white colonies (63 and  
20 100/ng vector for fractionated and unfractionated DNA, respectively). In this reaction, a ten-fold excess of reagents and enzymes were utilized to repair the sonicated DNA, which apparently improved the efficiency of cloning such molecules by two to three fold. The data collected from multiple cloning trials in Table 3, sections A and B, show that *CviJI*<sup>\*\*</sup> partial digestion results in three  
25 to sixteen times the number of white colonies than sonicated ER 1-treated DNA. Even with an optimal end repair reaction for the sonicated fragments, DNA treated with *CviJI*<sup>\*\*</sup> yielded three times more white colonies.

- 69 -

## Example 17

Analysis of *Cvi*JI\*\* Fragmentation for Shotgun Cloning and Sequencing

The ability of *Cvi*JI\*\* partial digestion to create uniformly representative clone libraries for DNA sequencing was tested on pUC19 DNA. pUC19 DNA was digested under *Cvi*JI\*\* conditions and size fractionated as described above. The fractionated DNA was cloned into the *Eco*RV site of M13SPSI, a lacZ minus vector constructed by adding an *Eco*RV restriction site to wild type M13 at position 5605. M13SPSI lacks a genetic cloning selection trait, therefore after ligation of the pUC19 fragments into the vector the sample was restricted with *Eco*RV to reduce the background of nonrecombinant plaques. Bacteriophage M13 plaques were picked at random and grown for 5-7 hours in 2 ml of 2XTY broth containing 20  $\mu$ l of a DH5 $\alpha$ F' overnight culture. After centrifugation to remove the cells, single-stranded phage DNA was purified using Sephaglass™ as specified by the manufacturer (Pharmacia LKB, Piscataway New Jersey). The single-stranded DNA was sequenced by the dideoxy chain termination method using a radiolabeled M13-specific primer and *Bst* DNA polymerase (Mead *et al.*, *Biotechniques* 11:76-87 (1991)). The first 100 bases of 76 randomly chosen clones were sequenced to determine which *Cvi*JI recognition site was utilized, the orientation of each insert and how effectively the cloned fragments covered the entire molecule, as shown in Figure 11. The positions of the 45 normal *Cvi*JI sites (PuGCPy) in pUC19 are indicated beneath the line labeled "NORMAL" in the Figure 11. Similarly, the 160 *Cvi*JI\* sites (GC) are indicated beneath the line labeled "RELAXED" in Figure 11. The marks above these lines indicate the *Cvi*JI\*\* pUC19 sites which were found in the set of 76 sequenced random clones. The frequency of cloning a particular site is indicated by the height of the line, and the left or right orientation of each clone is also indicated at the top of each mark. There are a total of 205 *Cvi*JI and *Cvi*JI\* sites in pUC19.



- 70 -

The data presented in Figure 11 demonstrate that, under CviJI<sup>\*\*</sup> partial conditions, normal CviJI sites are preferentially restricted over relaxed (CviJI<sup>\*</sup>) sites. Of the 76 clones that were analyzed, only 13%, or 1 in 7, had sequence junctions corresponding to a relaxed CviJI<sup>\*</sup> site. Thirty-five of the  
5 forty-five possible normal restriction sites were cloned, as compared to eight of the possible one hundred sixty relaxed sites. If the enzyme had exhibited no preference for normal or relaxed sites under the CviJI<sup>\*\*</sup> partial conditions utilized here, then 78% of the sequence junctions analyzed should have been generated by cleavage at a relaxed CviJI<sup>\*</sup> site. It may be noted that the relaxed CviJI<sup>\*</sup>  
10 restriction sites that were found appear to be clustered in two regions of the plasmid that are deficient in normal CviJI sites. In addition, the combined distribution of the normal and relaxed sites which were restricted to generate the 76 clones appears to be quasi-random. That is, the longest gap between cloned restriction sites was no greater than 250 bp and no one particular site is over-  
15 utilized.

A detailed analysis of the distribution of CviJI<sup>\*\*</sup> sequence junctions found from cloning pUC19 is presented in Table 5.

TABLE 5  
Distribution of Cloned CviJI\*\* Partially-Digested pUC19 sites.

Classification Group	Recognition Sequence	NGCN		Cloned CviJI** Distribution (%)	Pu/Py Structure
		Site Distribution in pUC19 (%)			
Normal (N)	A C	AGCC 9	(4.4)	13 (17.1)	PuPuPyPy
	GC	GGCC 11	(5.4)	16 (21.1)	
	G T	GGCT 10	(4.9)	12 (15.8)	
		AGCT 15	(7.3)	25 (32.9)	
		45	(22.0)	66 (86.9)	
Relaxed (R <sub>1</sub> )	C C	CGCC 11	(5.4)	0	PyPuPyPy
	GC	TGCC 12	(5.9)	2 (2.6)	
	T T	TGCT 10	(4.9)	1 (1.3)	
		CGCT 22	(10.7)	2 (2.6)	
		55	(26.9)	5 (6.5)	

TABLE 5 (cont'd)

Classification Group	Recognition Sequence	NGCN		Cloned CviJI** Distribution (%)	Pu/Py Structure
		Site Distribution in pUC19 (%)			
Relaxed (R <sub>2</sub> )	A A	AGCA 16	(7.3)	1 (1.3)	PuPuPyPu
	GC	GGCA 8	(3.9)	0	
	G G	AGCG 11	(5.4)	0	
		GGCG 22	(10.7)	4 (5.2)	
		57	(27.8)	5 (6.6)	
Relaxed (R <sub>3</sub> )	C A	CGCA 10	(4.9)	0	PyPuPyPu
	GC	TGCA 13	(6.3)	0	
	T G	CGCG 10	(4.9)	0	
		TGCG 15	(7.3)	0	
		48	(23.4)	0	

- 73 -

The GC sites in pUC19 may be divided into four classes based on their flanking Pu/Py structure. The fraction of GC sites observed in pUC19 which belong to each classification is roughly equal (22.0-27.8%). A striking difference was found between the observed distribution in pUC19 of normal and relaxed (R1, R2, R3) CviJI<sup>\*\*\*</sup> recognition sites and the distribution revealed by shotgun cloning and sequence analysis of CviJI<sup>\*\*\*</sup>-treated DNA. While most of the sites cleaved by this treatment were found to be PuGCPy (about 87%), or "normal" restriction sites, a significant fraction of the cleavage occurred at PyGCPy (about 6.5%) and PuGCPu (about 6.6%) sites, considering the short incubation times and limiting enzyme concentrations. The latter two categories of sites, and presumably the PyGCPu sites as well, are completely restricted under "relaxed" conditions, provided an excess of enzyme is present and sufficient time is allowed (see Figure 7, and Xia *et al.*, *Nucleic Acids Res.* 15:6075-6090 (1987)).

Digestion using CviJI<sup>\*\*\*</sup> treatment results in a relatively even distribution of breakage points across the length of the molecule (as shown in Figure 11). As described above, Figure 11 depicts a linear map of pUC19 showing the relative position of the lacZ' gene ( $\alpha$  peptide of  $\beta$ -galactosidase gene) and ampicillin resistance gene (Amp). The marks extending beneath the top line (labeled "NORMAL") show the relative position of the 45 normal CviJI sites (PuGCPy) present in pUC19. The marks above the line are the cleavage sites found from sequencing the CviJI<sup>\*\*\*</sup> partial library. The height of the line indicates the number of clones obtained from cleavage at that site, and the orientation of the flag designates the right or left orientation of the respective clone. The marks extending beneath the second line (labeled "RELAXED") show the relative positions of the 160 CviJI<sup>\*</sup> sites (GC) present in pUC19. Those marks above the line were found from sequencing the CviJI<sup>\*\*\*</sup> partial library. The bottom portion of Figure 11 shows the relative position and orientation of the first 20 clones sequenced, assuming a 350 bp read per clone. CviJI<sup>\*\*\*</sup> cleavage at relaxed sites appears to be important in "filling gaps" left by normal restriction.

- 74 -

The primary goal of this effort was to determine the efficacy of these methods for rapid shotgun cloning and sequencing. For these purposes, only 100 bases of sequence data were acquired per clone. However, if 350 bases of sequence had been determined from each clone, then the entire sequence of pUC19 would have been assembled from the overlap of the first 20 clones (Figure 11). In this sequencing simulation 75 % of pUC19 would have been sequenced at least 2 times from the first 20 clones. The highest degree of overfold sequencing would have been 6, and only involved 2.2% of the DNA. Figure 11 also shows that most of the 1x sequencing coverage occurred in a region of the plasmid with a very low density of normal and relaxed *Cvi*JI restriction sites. Most of the single coverage occurs in a 240 bp region of the plasmid between 1490 bp and 1730 bp where there are only 4 *Cvi*JI relaxed sites. It should also be noted that by the 27th randomly picked clone most of this region would have been covered a second time.

Shotgun sequencing strategies are efficient for accumulating the first 80-95% of the sequence data. However, the random nature of the method means that the rate at which new sequence is accumulated decreases as more clones are analyzed. In Figure 12 the total amount of unique pUC19 sequence accumulated was plotted as a function of the number of clones sequenced. The points represent a plot of the total amount of determined pUC19 sequence versus the total number of clones sequenced. The horizontal dashed line demarcates the 2686 bp length of pUC19. The smooth curve represents a continuous plot of the discrete function  $S(N) = NLe^{-cS} [((e^{cS}-1)/c) + (1-s)]$ . The theoretical accumulation curve expected for a process in which sequence information is acquired in a totally random fashion is also shown. The smooth curve is a continuous plot of the discrete function  $S(N)$  where

$$S(N) = NLe^{-c\sigma} [((e^{c\sigma}-1)/c) + (1-\sigma)].$$

This equation is based upon the results developed by Lander *et al.*, *Genomics* 2:231-239 (1988) for the progress of contig generation in genetic mapping. In the

- 75 -

equation:  $N$  is the number of clones sequenced,  $L$  is the length of clone insert in bp,  $c$  is the redundancy of coverage or  $LN/G$  (where  $G$  is length of fragment being sequenced in bp), and  $\sigma = 1 - \Theta$ , where  $\Theta$  is the fraction of length that two clones must share. The curve in Figure 12 was calculated with  $G = 2686$  bp,  $L = 350$  bp, and  $\sigma = 1$ . The plotted points lie close to the theoretical curve, and it thus appears that the sequence of pUC19 was accumulated in an apparent random fashion utilizing  $CviJI^{***}$  fragmentation and column fractionation.

### Example 18

#### Shotgun Cloning Utilizing 200 ng of Lambda DNA

Generally, 2-5  $\mu$ g of DNA are needed for the sonication and agarose gel fractionation method of shotgun cloning in order to provide the several hundred colonies or plaques required for sequence analysis (Bankier *et al. Methods in Enzymol.* 155:51-93 (1987)). A ten-fold reduction in the amount of substrate required greatly simplifies the construction of such libraries, especially from large genomes, (Davidson, *J. DNA Sequencing and Mapping* 1:389-394 (1991)). The efficiency of constructing a large shotgun library from nanogram amounts of substrate was tested utilizing 200 ng of  $CviJI^{***}$ -digested lambda DNA. This material was column-fractionated as described previously. In this case, 1/2 of the column eluant (15  $\mu$ l containing 50 ng of DNA) was ligated to 100 ng of *HincII*-digested and dephosphorylated pUC19 as described in Example 15. The cloning efficiencies of the control DNAs were similar to those reported in Tables 2 and 3. The 50 ng cloning experiment yielded 230 white colonies per ligation reaction in one trial, and 410 white colonies per ligation reaction in a second trial. Thus, it should be possible to routinely construct useful quasi-random shotgun libraries from as little as 0.2 - 0.5  $\mu$ g of starting material.

- 76 -

### Example 19

#### Epitope Mapping

*Cvi*JI\* recognizes the sequence GC (except for PyGCPu) in the target DNA. Under partial restriction conditions the length of fragment may be controlled by incubation time. Epitope mapping using *Cvi*JI\*\*\* partial digests involves generating DNA fragments of 100-300 bp from a cDNA coding for the protein of interest, by methods described in Example 13, inserting them into an M13 expression vector, plating out on solid media, lifting plaques onto a membrane, screening for binding to the ligand of interest, and picking the positive plaques for isolation of the DNA, which is then sequenced to identify the epitope. Thus, the same epitope may be expressed as a small fragment or a larger fragment. This approach allows one to determine the smallest fragment containing the epitope of interest using functional assays such as binding to an antibody or other ligand, or using a direct assay for activity. For insertion into an M13 vector, linkers may be added to the fragments or the insert may be dephosphorylated to ensure that each fragment is cloned alone without ligation of multiple inserts.

The expression vectors recommended for subcloning of the *Cvi*JI fragments are Lambda Zap™ (Stratagene, LaJolla, California) or bacteriophage M13-epitope display vectors. An advantage of using an M13-based vector is that the peptide or protein of interest may be displayed along with the M13 coat protein and does not require host cell lysis in order to analyze the protein of interest. The lambda-based vectors yield plaques and hence the protein can be directly bound to a membrane filter.

- 77 -

## Example 20

## CGase I

CGase I as used herein, refers to a restriction endonuclease reagent which cleaves DNA at the dinucleotide CG. CGase I activity is based on the combined  
5 star activities of the restriction endonucleases Hpa II and Taq I. Under normal reaction conditions (10 mM Bis Tris Propane-HCl pH 7.0, 10 mM MgCl<sub>2</sub>, 1 mM DTT; 1 unit of enzyme/ $\mu$ g DNA, 37°C for 1 hr), Hpa II recognizes CCGG and cleaves after the first C to leave a 2-base 5' overhang. Under normal reaction  
10 conditions (100 mM NaCl, 10 mM Tris-HCl pH 8.4, 10 mM MgCl<sub>2</sub>, 10 mM 2-mercaptoethanol, 1 unit of enzyme/ $\mu$ g DNA, 65°C for 1 hr) the restriction endonuclease Taq I recognizes TCGA and cleaves after the T to leave a 2-base 5' overhang.

Reaction conditions have been described for Taq I\* activity which decrease the cleavage specificity of Taq I (10 mM Tris-HCl pH 9.0, 5 mM MgCl<sub>2</sub>, 6 mM  
15 2-mercaptoethanol, 20% DMSO; 2000 units of enzyme/ $\mu$ g DNA, 65°C for 1 hr) (Barany, Gene, 65:149-165 (1988)). These reaction conditions allow Taq I\* to cleave DNA at the following sequences:

20 Taq I\*            TCGA  
                 CCGA (TCGG)  
                 ACGA (TCGT)  
                 TCTA (TAGA)  
                 TCAA (TTGA)  
                 GCGA (TCGC)

We are unaware of any literature descriptions of Hpa II\* conditions.  
25 However, the following conditions were established to promote Hpa II\* activity which are also compatible with Taq I\* activity: 5 mM KCl, 10 mM Tris-HCl pH 8.5, 10 mM MgCl<sub>2</sub>, 1 mM DTT, 15% DMSO, 100  $\mu$ g/ml BSA (CGase buffer); 50 units of enzyme/ $\mu$ g DNA 50°C for 1 hr. The Hpa II\* recognition sites were



- 78 -

determined by cloning and sequencing Hpa II\* restricted fragments. The characterized Hpa II\* recognition sequences are as follows:

5 Hpa II\* CCGG  
CCGC (GCGG)  
CCGA (TCGG)  
ACGG (CCGT)

10 Taq I (400 units/ $\mu$ g DNA) and Hpa II (50 units/ $\mu$ g DNA) were then combined (CGase I) in CGase I buffer and the following recognition sites were identified by cloning and sequencing restricted pUC19 fragments.

15 CGase I GCGC  
TCGA  
CCGG  
GCGT  
ACGA  
ACGG (CCGT)  
GCGG (CCGC)  
CCGA (TCGG)

20 CGase I restriction of natural DNA, (i.e. pUC19, lambda), results in fragments ranging from 20-200 bp in length (average 20-60 bp). Heat denaturation of these fragments generates numerous oligonucleotides of variable length but precise specificity for the cognate template as was the case with CviJ I\* digestion. CGase I restriction of the small plasmid pUC19 (2689 bp) theoretically yields 174 restriction fragments, or 384 oligonucleotides after a heat denaturation step.

25 The "two-cutter" activity of CviJ I\* and CGase I represent a unique class of restriction endonuclease activity in that no other known restriction endonucleases will generate this size range of oligonucleotides. The ability to generate numerous oligonucleotides with perfect sequence specificity from any DNA, without regard to sequence composition, genetic origin, or prior sequence  
30 knowledge is one of the properties that CGase I shares with CviJ I\*. In addition, the generation of numerous oligonucleotides by CviJ I or CGase I results in a

- 79 -

form of probe or primer amplification not practical using conventional means of organic synthesis.

Based on ability to recognize a dinucleotide sequence, the present invention contemplates the interchangeability of CGase I with CviJ I\* in all of the applications described herein.

5

### Example 21

#### Purification of CviJ I Restriction Endonuclease from IL-3A-Infected *Chlorella* Cells

CviJ I was prepared by a modification of the method described by Xia *et al.*, *Nucl. Acids Res.* 15:6025-6090 (1987). *Chlorella* NC64A cells (ATCC Accession No. 75399 deposited on January 21, 1993, American Type Culture Collection, Rockville, Maryland) were infected with the virus IL-3A (ATCC Accession No. 75354 deposited November 6, 1992, American Type Culture Collection, Rockville, Maryland) according to Van Etten *et al.*, *Virology* 126:117-125 (1983). Five grams of IL-3A infected *Chlorella* NC64A cells were suspended in a glass homogenization flask with 15 g of 0.3 mm glass beads in buffer A (10 mM Tris-HCl pH 7.9, 10 mM 2-mercaptoethanol, 50 µg/ml phenylmethylsulfonyl fluoride (PMSF), 20 µg/ml benzamidine, 2 µg/ml o-phenanthroline). Cell lysis was carried out at 4000 rpm for 90 sec in a Braun MSK mechanical homogenizer (Allentown, PA) with cooling from a CO<sub>2</sub> tank. After lysis 2 M NaCl was added to a final concentration of 200 mM, after which 10% polyethyleneimine (PEI) (Life Technologies, Bethesda, MD) (pH 7.5) was added to a final concentration of 0.3%. The mixture was then stirred for 2 hrs. at 4°C then centrifuged for 1 hr. at 50,000 g. Ammonium sulfate was added to the supernatant to 70% saturation and stirred overnight. A protein pellet was recovered by centrifugation for 1 hr. at 50,000 g. The resulting pellet was dissolved in 20 ml of buffer B (20 mM Tris-acetate pH 7.5, 0.5 mM EDTA, 10

10

15

20

25

- 80 -

mM 2-mercaptoethanol, 10% glycerol, 30 mM KCl, 50 ug/ml PMSF, 20 µg/ml benzamidine [Sigma, St. Louis, Missouri], 2 µg/ml o-phenanthroline [Sigma]) and dialysed against 500 ml of buffer B with 3 changes. The dialysed solution was then applied to 1 x 6 cm Heparin-Sepharose (Pharmacia LKB, Piscataway, New Jersey) column. After a 50 ml wash with buffer B, a 100 ml gradient of 0 to 0.7 M KCl in buffer B was run. Fractions having CviJ I activity as measured by digestion of pUC19 DNA and agarose gel electrophoresis, were pooled, diluted in 5 volumes of buffer C (10 mM K/PO<sub>4</sub> pH 7.4, 0.5 mM EDTA, 10 mM 2-mercaptoethanol, 75 mM NaCl, 0.05% Triton X-100, 10% glycerol, 50 µg/ml PMSF, 20 µg/ml benzamidine, 2 µg/ml o-phenanthroline) and applied to a 1 x 7 cm Phosphocellulose P11 (Whatman) column equilibrated in buffer C. After washing with 30 ml of buffer C, CviJ I was eluted by a 100 ml gradient of 0 to 0.7 M NaCl in buffer C. At this step CviJ I activity separated from non-specific nucleases. CviJ I containing fractions were pooled and diluted in 4 volumes of buffer C and applied to a 1 x 4 cm hydroxyapatite HTP column (BioRad, Hercules, CA). After washing with 30 ml of buffer C, CviJ I was eluted by a 0 to 0.7 M potassium phosphate (pH 7.4) gradient in buffer C. Active fractions containing CviJ I activity and lacking non-specific nuclease activity were pooled and were dialysed overnight against storage buffer (50 mM potassium phosphate 200 mM KCl, 0.5 mM EDTA, 50% glycerol, 20 ug/ml PMSF were pooled) and stored at -20°C.

Although the present invention has been described in types of preferred embodiments, it is intended that the present invention encompass all modifications and variations which occur to those skilled in the art upon consideration of the disclosure herein, and in particular those embodiments which are within the broadest proper interpretation of the claims and their requirements.

- 81 -

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: Molecular Biology Resources, Inc.
- (ii) TITLE OF INVENTION: Materials and Methods for  
Restriction Endonuclease Applications
- (iii) NUMBER OF SEQUENCES: 13
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
  - (B) STREET: 6300 Sears Tower, 233 South Wacker Drive
  - (C) CITY: Chicago
  - (D) STATE: Illinois
  - (E) COUNTRY: United States of America
  - (F) ZIP: 60606-6402
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER:
  - (B) FILING DATE:
  - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Clough, David W.
  - (B) REGISTRATION NUMBER: 36,107
  - (C) REFERENCE/DOCKET NUMBER: 28003/31967/PCT
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: 312/474-6300
  - (B) TELEFAX: 312/474-0448
  - (C) TELEX: 25-3856

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 44 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CAATTCACA CAGGAACAG CTATGTCTTT TCGCACGTTA GAAC

- 82 -

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5496 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

ATGTCTTTTC GCACGTTAGA ACTATTCGCC GGTATAGCTG GTATTCACA TGGCCTCAGA	60
GGTATATCTA CACCAGTTGC ATTCTAGAA ATTAATGAAG ACGCACAAAA ATTCTTGAAA	120
ACAAAGTTTT CAGATGCATC TGTATTCAAT GACGTTACGA AATTTACCAA ATCGGACTTC	180
CCAGAAGACA TAGACATGAT TACTGCGGGA TTCCCGTGCA CTGGGTTTAG TATTGCAGGT	240
TCTAGAAGTG GATTCTGAACA CAAGGAATCC GGTCTCTTTG CTGATGTTGT GCGAATCACC	300
GAAGAGTATA AACCTAAAAT AGTGTTTTTG GAAAACCTCC ATATGTTGTC CCACACTTAC	360
AATCTCGATG TCGTCGTAAA AAAGATGGAT GAAATTGGTT ATTTCTGCAA GTGGGTAAC	420
TGTCGGGCAT CAATTATAGG AGCCCATCAT CAACGCCACC GGTGGTTTTG TCTCGCGATT	480
CGAAAAGATT ATGAACCAGA AGAATAATT GTATCTGTGA ATGCTACAAA GTTCGACTGG	540
GAAAATAATG AACCACCGTG TCAAGTAGAC AATAAGAGTT ACGAGAATTC AACTCTTGTT	600
CGTCTGGCAG GATATTCCGT GGTCCCCGAC CAGATCAGAT ATGCTTTCAC CGGTCTATTT	660
ACAGGTGATT TTGAGTCATC GTGGAAAAC ACCTTGACAC CTGGGACAA ATTTGGCACC	720
GAACACAAAA AATGAAAGG AACTTACGAT AAAGTCATAA ACGGGTATTA TGAGAACCAT	780
GTGTATTATT CTTTTTCAAG GAAAGAAGTT CATCGCGCTC CTCTAAATAT ATCCGTGAAA	840
CCACGTGATA TTCCGGAGAA ACATAACGGA AAAACACTCG TAGATCGCGA AATGATCAAG	900
AAATATTGGT GCACACCATG TGCTAGTTAT GGCACGTGCTA CTGCTGGATG CAATGTTCTG	960
ACAGACCGTC AGTCACATGC ACTTCCTACA CAAGTCAGGT TTTCATATAG GGGTGTATGT	1020
GGACGACATT TGTCTGGTAT ATGGTGTGCA TGGTTGATGG GGTATGACCA AGAATATCTT	1080
GGTTATTTGG TTCAATATGA TTAAAATATT TTGATACACT AAATGGATAT AAGAAGAAAA	1140
CGTTTTACAA TAGAAGGGGC TAAACGTATA ATACTCGAAA AAAAGAGACT TGAAGAGAAA	1200
AAAAGAATTG CGGAAGAGAA AAAAAGAATT GCACCTTATAG AAAACAACG AATTGCGGAA	1260
GAGAAAAAAA GAATTGCGGA AGAGAAAAAA CGATTTCGAC TTGAAGAGAA AAAACGAATT	1320
GCGGAAGAAA AAAACGAAT CGCGGAAGAG AAAAAACGAA TCGTGGAAGA GAAAAAAGA	1380
CTTGCACTTA TAGAAAAACA ACGAATTGCG GAAGAGAAAA TTGCGTCGGG GAGAAAAATT	1440

- 83 -

AGAAAGAGGA TCTCTACAAA TGCAACAAAA CATGAAAGAG AATTTGTCAA AGTTATAAAT	1500
TCAATGTTTCG TCGGACCCGC TACTTTTGTA TTCGTAGATA TAAAAGGTAA TAAATCCAGA	1560
GAAATCCACA ACGTTGTAAG ATTCAGACAA TTACAAGGCA GTAAAGCGAA ATCCCCGACC	1620
GCGTATGTTG ATAGAGAATA TAACAAACCT AAAGCGGATA TAGCAGCGGT AGACATAACC	1680
GGTAAAGATG TGGCATGGAT ATCCCATAAA GCATCTGAAG GATATCAACA ATATCTAAAA	1740
ATTTCTGGAA AGAACCTCAA GTTCACAGGA AAAGAATTAG AAGAAGTTCT ATCGTTCAG	1800
AGAAAAGTAG TTAGTATGGC ACCGGTATCT AAAATATGGC CTGCTAATAA GACCGTATGG	1860
TCTCCTATCA AGTCAAATTT GATTAAAAAT CAAGCAATAT TCGGATTGA TTACGGTAAG	1920
AAACCAGGAA GGGACAATGT AGACATCATA GGTCAAGGAC GACCAATTAT AACAAAAAGA	1980
GGTTCCATAT TATATCTTAC ATTCACTGGT TTTAGCGCAT TAAATGGGCA CTTGGAGAAT	2040
TTTACTGGGA AACATGAACC CGTTTTCTAT GTAAGAACAG AACGGAGTAG TAGCGGGAGA	2100
AGTATAACAA CTGTCGTCAA TGGTGTCACT TATAAAATT TAAGATTCTT TATACATCCA	2160
TACAACTTTG TTTCTTCAAA AACACAACGT ATTATGTAGG ACCATTTTCC CGAGAGACTT	2220
TGTTGACCGC GTACTAAAAA ATGGTCACGA TATTTGTCTA AAGATGCTCA TAGAAGCAGG	2280
TGCAAACCTT GACATCGTCA GTGTTGAGTA TACACCATTA CATCTACATG TGGTGATATT	2340
TGTATAACG GTAAATACCT ATATATACAA TACGTATCCC CCTAAAAGCG CTTAGATTTT	2400
TTAGTTGTAT ACTACTTTTG TATAAGACCT GTAAGTTACA AACTAAAAGT TTCAGCTTTG	2460
CCTTCGAAAC AAGCAATTAC CGCATGAGAA TAATATCCAT TATGGATGTT TTCTGCTAAT	2520
AAAACGATAT TTCCTACAGA AGTTTCTATG ATTAGTCCG AAATATTGAG ATCATCGTCA	2580
CGTTTTTCTT TACCGTATTT TACTTTCGTG ATCGTCGCAC CAATAAAATC ATCTCGTGTG	2640
AGTTCATTG GCAATTGTGC CGTGACACCA AATCTCTCAC AACAACTTG ATGTCCATCC	2700
ATTGCTAACA CTATCGGTAA TCCATGTGTG GTGTGTACGA CCACACCGTT ATAACTATAA	2760
CACGTGTAGT TGTCGTCTAT ATCATATAAC TCGAGAGCGG TGTGAACTTC TTCAGATCTA	2820
TTATTAATCG GATCTGATCC ATAAGAAGAA TCTTCATATT TACAAATAAA ATCATCCGAT	2880
ATGTTCTGCA CACGAACAAC ATTCGTCAAA TTTCTGTGAT GACGAATCTC CATCTCTGAA	2940
TCATTAGAGA CTTGCGAGTA TATAACATTA TAATTGTTGA TATGATTATT ACGTTTCATA	3000
TCAACAAAAT ACATATAAAC ACCATACAAA TATTAACAAC CGTTAGTATA TAATGGATAA	3060
CATTTGCAAT AGTATATTCA CTGCAGTAAA AAATGGCCAC GAAGCTTGTT TGAAGATGAT	3120
GCTCATTGAA AGAGGTAGCA ATATCAATGA TGTTTCCGAA TCAAAATATG GAAATACACC	3180
ACTACATATT GCAGCTCATC ATGGTAATGA TGTGTGTTTG AAGATGCTTA TTGACGCAGG	3240
TGCAAACCTT GATATCACAG ATATTTCTGG AGGAACACCA CTTTCATCGT CCGTTTTGAA	3300

- 84 -

TGGCCATGAC	ATATTGTACA	GATGCTCGTA	GAAGCAGGTG	CAAACCTTAG	TATCATAACT	3360
AATTTGGGAT	GGATACCGTT	ACATTACGCG	GCTTTTAATG	GTAATGATGC	GATTTTGAGG	3420
ATGCTCATCG	TTGTAAGTGA	TAATGTTGAC	GTTATCAATG	ATCGCGGTTG	GACGGCGTTA	3480
CATTACGCGG	CTTTTAATGG	TCATAGCATG	TGCGTCAAGA	CGCTTATTGA	TGCGGGTGCA	3540
AATCTTGACA	TCACAGATAT	TTCGGGATGT	ACACCACTTC	ATCGTGCGGT	TTATAATGAC	3600
CACGATGCAT	GTGTGAAGAT	ACTCGTAGAA	GCAGGTGCAA	CTCTTGACGT	CATTGATGAT	3660
ACTGAGTGGG	TGCCGTTACA	TTACGCGGCT	TTTAATGGTA	ATGATGCGAT	TTTGAGGATG	3720
CTCATTGAAG	CAGGTGCAGA	TATTGATATA	TCTAATATAT	GTGATTGGAC	GGCGTTACAT	3780
TACGCGGCTC	GAAATGGACA	CGATGTGTGT	ATAAAAACAC	TCATCGAAGC	AGGTGGTAAC	3840
ATCAACGCCG	TCAACAAATC	GGGGGATACA	CCACTAGATA	TTGCAGCATG	TCATGACATT	3900
GCAGTATGTG	TGATCGTGAT	AGTCAATAAG	ATCGTTTCGG	AGCGGCCGTT	GCGTCCGAGT	3960
GAGTTGTGTG	TCATACCACC	AACGTCTGCT	GCATTAGGTG	ATGTGTTGCG	AACGACGATG	4020
CGGCTTCATG	GGCGATCGGA	AGCTGCAAAG	ATCACAGCGC	ATCTTCCTGT	GGGTGCAAGG	4080
GATACTCTAC	GAACTACTGC	GTTGTGTTTG	AACCGAACAA	TTCCGAGAG	ATCTCGTTGA	4140
TAGTGTATTA	ATTGAATGCG	TGTAAAGTTA	CGCTATTTTT	TTCCAAAAAG	GGTTTGCAATG	4200
AAATACAACA	CGATCTTTTG	TAGATCGTTT	ACCATTAGTT	GTATTCGTGC	AATAGAGACC	4260
ATACGTACCT	CCAAATTCAT	TTACTTTACC	TACAGTATTA	CCACTTCCTT	TTTTTCCTAT	4320
AGTAGTATCT	AAATTCAACC	CTTTGAACTC	ATGCCATTAT	ACAGACAGAG	CGTATGAACC	4380
GTTTTGTGCC	AATTTACCT	TCAAACGAT	AGTAACCCAT	TGACCTCTAG	GAATTTTAAC	4440
CGATCTTATA	AGTATCTGCT	TACTTCCAAG	TCCTTTTTCA	AAAGCATACA	ACGATCCTGT	4500
AAGGTTATCC	CCAGAACCCTG	AAATTGTAAA	GAACGACTGG	AAATGAATAG	GTTGCATTAG	4560
ATCTGTATAC	ATATCACTTG	GTTCGAAATG	AAAATCGTAG	TCCCAATTAG	GTACGTTCCA	4620
CCAAGTTTAA	TACGGGGTCT	TTCCACCGAG	ACCGGACATT	TCAGCACGAG	CCTTGTAAGA	4680
ATGATATGAT	GTGGTTAAAT	CTCTATCACC	ATCGTTCCAC	TTTCCTCTGA	ACCGAAGACC	4740
ATGCATCGTT	ATACCTGGTG	CAACCTGTAC	TAAATTCTTT	ATTTACAGGTG	CGGCTCCGGG	4800
TGGATTAACT	CGAGATTCGT	CAAATCTAAA	ATATGATAAC	GATGTTCCAA	CAGTAGAACC	4860
ACTGGGTGGT	ATGGCAGTTG	CTGGAAGGGA	AGGTAAAAC	TTAGGATATT	TCAAATCACC	4920
AACACCTTGA	GGGTTTACTT	GAATACTTCT	GGGAGATGTT	GGTGGTTTCG	TCGAAGGTGG	4980
TTTCGTTGAA	GGTGGTTTCG	TCGAAGGTGG	TTTCGTCGAA	GGTGGTTTCG	TCGAAGGTGG	5040
TTTCGTCGAA	GGTGGTTTCG	TCGAAGGTGG	TTTCGTCGAA	GGTGGTTTCG	TCGAAGGTGG	5100
TTTCGTCGAA	GGTGGTTTCG	TCGAAGGTGG	TTTCGTCGAA	GGTGGTTTCG	TCGAAGGTGG	5160

- 85 -

TTTCGTCGAA GGTGGTTTCG TCGAAGGTGG TTTCGTTGGC GGAAGTGGGG CATGACCATA 5220  
 ATCCGTTAAA TTCCCGCATT CACCTAATGA TGTACTCCAT AAAGAACCGG GTGCGCATTG 5280  
 CATTCTTATT GGTTCGTAG TATCAGATAT ACATACGAAA TAATGAGAAT CATTTCCCT 5340  
 GCCAAATAAT TTACCAGATT TGCCTTTACA TGACATTATT TGTAAATATA TATTATTATA 5400  
 ATTTTAAAAA AACTAACGTC TATTTAAAAT TATGTAATAC GTATTATATC AATGCATCAT 5460  
 CTTAATCATT TCCTAACGTA TAAGCGTAGC GAATTC 5496

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1225 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: join(1..33, 55..1128)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CAA GAA TAT CTT GGT TAT TTG GTT CAA TAT GAT TAAAAATATTT TGATACACTA 53  
 Gln Glu Tyr Leu Gly Tyr Leu Val Gln Tyr Asp  
 1 5 10

A ATG GAT ATA AGA AGA AAA CGT TTT ACA ATA GAA GGG GCT AAA CGT 99  
 Met Asp Ile Arg Arg Lys Arg Phe Thr Ile Glu Gly Ala Lys Arg  
 15 20 25

ATA ATA CTC GAA AAA AAG AGA CTT GAA GAG AAA AAA AGA ATT GCG GAA 147  
 Ile Ile Leu Glu Lys Lys Arg Leu Glu Glu Lys Lys Arg Ile Ala Glu  
 30 35 40

GAG AAA AAA AGA ATT GCA CTT ATA GAA AAA CAA CGA ATT GCG GAA GAG 195  
 Glu Lys Lys Arg Ile Ala Leu Ile Glu Lys Gln Arg Ile Ala Glu Glu  
 45 50 55

AAA AAA AGA ATT GCG GAA GAG AAA AAA CGA TTC GCA CTT GAA GAG AAA 243  
 Lys Lys Arg Ile Ala Glu Glu Lys Lys Arg Phe Ala Leu Glu Glu Lys  
 60 65 70

AAA CGA ATT GCG GAA GAA AAA AAA CGA ATC GCG GAA GAG AAA AAA CGA 291  
 Lys Arg Ile Ala Glu Glu Lys Lys Arg Ile Ala Glu Glu Lys Lys Arg  
 75 80 85 90

ATC GTG GAA GAG AAA AAA AGA CTT GCA CTT ATA GAA AAA CAA CGA ATT 339  
 Ile Val Glu Glu Lys Lys Arg Leu Ala Leu Ile Glu Lys Gln Arg Ile  
 95 100 105

GCG GAA GAG AAA ATT GCG TCG GGG AGA AAA ATT AGA AAG AGG ATC TCT 387  
 Ala Glu Glu Lys Ile Ala Ser Gly Arg Lys Ile Arg Lys Arg Ile Ser  
 110 115 120



- 86 -

ACA AAT GCA ACA AAA CAT GAA AGA GAA TTT GTC AAA GTT ATA AAT TCA Thr Asn Ala Thr Lys His Glu Arg Glu Phe Val Lys Val Ile Asn Ser 125 130 135	435
ATG TTC GTC GGA CCC GCT ACT TTT GTA TTC GTA GAT ATA AAA GGT AAT Met Phe Val Gly Pro Ala Thr Phe Val Phe Val Asp Ile Lys Gly Asn 140 145 150	483
AAA TCC AGA GAA ATC CAC AAC GTT GTA AGA TTC AGA CAA TTA CAA GGC Lys Ser Arg Glu Ile His Asn Val Val Arg Phe Arg Gln Leu Gln Gly 155 160 165 170	531
AGT AAA GCG AAA TCC CCG ACC GCG TAT GTT GAT AGA GAA TAT AAC AAA Ser Lys Ala Lys Ser Pro Thr Ala Tyr Val Asp Arg Glu Tyr Asn Lys 175 180 185	579
CCT AAA GCG GAT ATA GCA GCG GTA GAC ATA ACC GGT AAA GAT GTG GCA Pro Lys Ala Asp Ile Ala Ala Val Asp Ile Thr Gly Lys Asp Val Ala 190 195 200	627
TGG ATA TCC CAT AAA GCA TCT GAA GGA TAT CAA CAA TAT CTA AAA ATT Trp Ile Ser His Lys Ala Ser Glu Gly Tyr Gln Gln Tyr Leu Lys Ile 205 210 215	675
TCT GGA AAG AAC CTC AAG TTC ACA GGA AAA GAA TTA GAA GAA GTT CTA Ser Gly Lys Asn Leu Lys Phe Thr Gly Lys Glu Leu Glu Glu Val Leu 220 225 230	723
TCG TTC AAG AGA AAA GTA GTT AGT ATG GCA CCG GTA TCT AAA ATA TGG Ser Phe Lys Arg Lys Val Val Ser Met Ala Pro Val Ser Lys Ile Trp 235 240 245 250	771
CCT GCT AAT AAG ACC GTA TGG TCT CCT ATC AAG TCA AAT TTG ATT AAA Pro Ala Asn Lys Thr Val Trp Ser Pro Ile Lys Ser Asn Leu Ile Lys 255 260 265	819
AAT CAA GCA ATA TTC GGA TTT GAT TAC GGT AAG AAA CCA GGA AGG GAC Asn Gln Ala Ile Phe Gly Phe Asp Tyr Gly Lys Lys Pro Gly Arg Asp 270 275 280	867
AAT GTA GAC ATC ATA GGT CAA GGA CGA CCA ATT ATA ACA AAA AGA GGT Asn Val Asp Ile Ile Gly Gln Gly Arg Pro Ile Ile Thr Lys Arg Gly 285 290 295	915
TCC ATA TTA TAT CTT ACA TTC ACT GGT TTT AGC GCA TTA AAT GGG CAC Ser Ile Leu Tyr Leu Thr Phe Thr Gly Phe Ser Ala Leu Asn Gly His 300 305 310	963
TTG GAG AAT TTT ACT GGG AAA CAT GAA CCC GTT TTC TAT GTA AGA ACA Leu Glu Asn Phe Thr Gly Lys His Glu Pro Val Phe Tyr Val Arg Thr 315 320 325 330	1011
GAA CGG AGT AGT AGC GGG AGA AGT ATA ACA ACT GTC GTC AAT GGT GTC Glu Arg Ser Ser Ser Gly Arg Ser Ile Thr Thr Val Val Asn Gly Val 335 340 345	1059
ACT TAT AAA AAT TTA AGA TTC TTT ATA CAT CCA TAC AAC TTT GTT TCT Thr Tyr Lys Asn Leu Arg Phe Phe Ile His Pro Tyr Asn Phe Val Ser 350 355 360	1107

- 87 -

TCA AAA ACA CAA CGT ATT ATG TAGGACCATT TTCCCGAGAG ACTTTGTTGA 1158  
 Ser Lys Thr Gln Arg Ile Met  
 365

CCGCGTACTA AAAAATGGTC ACGATATTTG TCTAAAGATG CTCATAGAAG CAGGTGCAAA 1218  
 CCTTGAC 1225

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 369 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Gln Glu Tyr Leu Gly Tyr Leu Val Gln Tyr Asp Met Asp Ile Arg Arg  
 1 5 10 15  
 Lys Arg Phe Thr Ile Glu Gly Ala Lys Arg Ile Ile Leu Glu Lys Lys  
 20 25 30  
 Arg Leu Glu Glu Lys Lys Arg Ile Ala Glu Glu Lys Lys Arg Ile Ala  
 35 40 45  
 Leu Ile Glu Lys Gln Arg Ile Ala Glu Glu Lys Lys Arg Ile Ala Glu  
 50 55 60  
 Glu Lys Lys Arg Phe Ala Leu Glu Glu Lys Lys Arg Ile Ala Glu Glu  
 65 70 75 80  
 Lys Lys Arg Ile Ala Glu Glu Lys Lys Arg Ile Val Glu Glu Lys Lys  
 85 90 95  
 Arg Leu Ala Leu Ile Glu Lys Gln Arg Ile Ala Glu Glu Lys Ile Ala  
 100 105 110  
 Ser Gly Arg Lys Ile Arg Lys Arg Ile Ser Thr Asn Ala Thr Lys His  
 115 120 125  
 Glu Arg Glu Phe Val Lys Val Ile Asn Ser Met Phe Val Gly Pro Ala  
 130 135 140  
 Thr Phe Val Phe Val Asp Ile Lys Gly Asn Lys Ser Arg Glu Ile His  
 145 150 155 160  
 Asn Val Val Arg Phe Arg Gln Leu Gln Gly Ser Lys Ala Lys Ser Pro  
 165 170 175  
 Thr Ala Tyr Val Asp Arg Glu Tyr Asn Lys Pro Lys Ala Asp Ile Ala  
 180 185 190  
 Ala Val Asp Ile Thr Gly Lys Asp Val Ala Trp Ile Ser His Lys Ala  
 195 200 205  
 Ser Glu Gly Tyr Gln Gln Tyr Leu Lys Ile Ser Gly Lys Asn Leu Lys  
 210 215 220

- 88 -

Phe Thr Gly Lys Glu Leu Glu Glu Val Leu Ser Phe Lys Arg Lys Val  
 225 230 235 240  
 Val Ser Met Ala Pro Val Ser Lys Ile Trp Pro Ala Asn Lys Thr Val  
 245 250 255  
 Trp Ser Pro Ile Lys Ser Asn Leu Ile Lys Asn Gln Ala Ile Phe Gly  
 260 265 270  
 Phe Asp Tyr Gly Lys Lys Pro Gly Arg Asp Asn Val Asp Ile Ile Gly  
 275 280 285  
 Gln Gly Arg Pro Ile Ile Thr Lys Arg Gly Ser Ile Leu Tyr Leu Thr  
 290 295 300  
 Phe Thr Gly Phe Ser Ala Leu Asn Gly His Leu Glu Asn Phe Thr Gly  
 305 310 315 320  
 Lys His Glu Pro Val Phe Tyr Val Arg Thr Glu Arg Ser Ser Ser Gly  
 325 330 335  
 Arg Ser Ile Thr Thr Val Val Asn Gly Val Thr Tyr Lys Asn Leu Arg  
 340 345 350  
 Phe Phe Ile His Pro Tyr Asn Phe Val Ser Ser Lys Thr Gln Arg Ile  
 355 360 365  
 Met

## (2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 17 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GTAAAACGAC GGCCAGT

17

## (2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 16 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GCCAAGCTTC GATGAT

16

- 89 -

## (2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 31 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ATCTTCGCGA ATCACTGGC CGTCGTTT C

31

## (2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 14 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GAATTCGCGA AGAT

14

## (2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 33 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ATCATCCAAG CTTGGCACTG GCCGTCGTTT TAC

33

## (2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 81 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GTAAACGAC GGCCAGTGAA TTCGCGAAGA TNNNNNNNNN NNNNNNNNAT CATCCAAGCT

60

TGGCACTGGC CGTCGTTT C

81

(2) INFORMATION FOR SEQ ID NO:11:

- (ii) MOLECULE TYPE: DNA

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GTAAACGAC	GGCCAGTGCC	AAGCTTGGAT	GATNNNNNNN	NNNNNNNNNN	ATCTTCGCGA	60
ATTCAGTGGC	CGTCGTTTTA	C				81

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 270 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)

- (ix) FEATURE:**

- ```
(A) NAME/KEY: CDS
(B) LOCATION: join(26..148, 190..207, 244..270)
```

- (xi) SEQUENCE DESCRIPTION: SEO ID NO:12:

|                                                                      |                                                |  |     |
|----------------------------------------------------------------------|------------------------------------------------|--|-----|
| TAACAATTTC ACACAGGAAA CAGCT                                          | ATG ACC ATG ATT ACG CCA AGC TCG AAA            |  | 52  |
|                                                                      | Met Thr Met Ile Thr Pro Ser Ser Lys            |  |     |
|                                                                      | 1                    5                         |  |     |
| TTA ACC CTC ACT AAA GGG AAC AAA AGC TGG TAC CGG GGC CCC CCC TCG      |                                                |  | 100 |
| Leu Thr Leu Thr Lys Gly Asn Lys Ser Trp Tyr Arg Gly Pro Pro Ser      |                                                |  |     |
| 10                    15                    20                    25 |                                                |  |     |
| AGG TCG ACG GTA TCG ATA AGC TTG ATA AAC CAT TTA TAC AAT AAG CGT      |                                                |  | 148 |
| Arg Ser Thr Val Ser Ile Ser Leu Ile Asn His Leu Tyr Asn Lys Arg      |                                                |  |     |
|                                                                      | 30                    35                    40 |  |     |
| TGATATAAGT TTGTATATAC GTCATTTCTGT TATATCAACA A ATG TTA TCA TAT       |                                                |  | 201 |
|                                                                      | Met Leu Ser Tyr                                |  |     |
|                                                                      | 45                                             |  |     |
| TAT ACG TAAACTGGC TTAAAAA AAA ACAGAGTGTA ACTATA ATG TCT TTT CGC      |                                                |  | 255 |
| Tyr Thr                                                              | Met Ser Phe Arg                                |  |     |
|                                                                      | 50                                             |  |     |
| ACG TTA GAA CTA TTT                                                  |                                                |  | 270 |
| Thr Leu Glu Leu Phe                                                  |                                                |  |     |
| 55                                                                   |                                                |  |     |

- 91 -

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 56 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

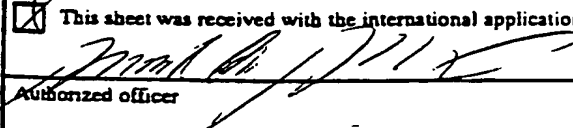
(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
Met Thr Met Ile Thr Pro Ser Ser Lys Leu Thr Leu Thr Lys Gly Asn  
1      5      10     15  
Lys Ser Trp Tyr Arg Gly Pro Pro Ser Arg Ser Thr Val Ser Ile Ser  
20    25    30  
Leu Ile Asn His Leu Tyr Asn Lys Arg Met Leu Ser Tyr Tyr Thr Met  
35    40    45  
Ser Phe Arg Thr Leu Glu Leu Phe  
50    55
```

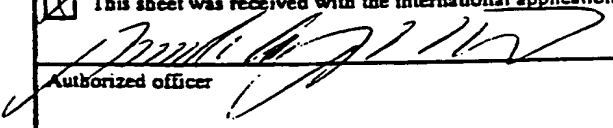
## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

|                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| A. The indications made below relate to the microorganism referred to in the description<br>on page <u>79</u> , line <u>13</u>                                                                                                                                                                                                                                                                                                                 |                                                                                                                                                                                     |
| B. IDENTIFICATION OF DEPOSIT <span style="float: right;">Further deposits are identified on an additional sheet <input checked="" type="checkbox"/></span>                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                     |
| Name of depositary institution<br><br>American Type Culture Collection                                                                                                                                                                                                                                                                                                                                                                         |                                                                                                                                                                                     |
| Address of depositary institution (including postal code and country)<br><br>12301 Parklawn Drive<br>Rockville, Maryland 20852<br>UNITED STATES OF AMERICA                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                     |
| Date of deposit<br><br>November 6, 1992                                                                                                                                                                                                                                                                                                                                                                                                        | Accession Number<br><br>A.T.C.C. 75354                                                                                                                                              |
| C. ADDITIONAL INDICATIONS (leave blank if not applicable) <span style="float: right;">This information is continued on an additional sheet <input type="checkbox"/></span>                                                                                                                                                                                                                                                                     |                                                                                                                                                                                     |
| <p>"In respect of those designations in which a European patent is sought, a sample of the deposited microorganism will be made available until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample (Rule 23(4) EPC)."</p> |                                                                                                                                                                                     |
| D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                     |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                     |
| E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                                                                                     |
| The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")                                                                                                                                                                                                                                                                           |                                                                                                                                                                                     |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                     |
| <div>For receiving Office use only</div> <div><input checked="" type="checkbox"/> This sheet was received with the international application</div> <div></div> <div>Authorized officer</div>                                                                                                                                                                | <div>For International Bureau use only</div> <div><input type="checkbox"/> This sheet was received by the International Bureau on:</div> <div> </div> <div>Authorized officer</div> |

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

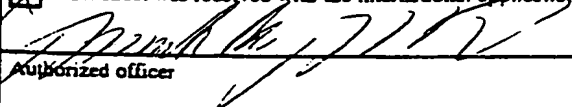
(PCT Rule 13bis)

|                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                        |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| A. The indications made below relate to the microorganism referred to in the description<br>on page <u>79</u> , line <u>10</u>                                                                                                                                                                                                                                                                                                                 |                                                                                                                                                                        |
| B. IDENTIFICATION OF DEPOSIT <span style="float: right;">Further deposits are identified on an additional sheet <input checked="" type="checkbox"/></span>                                                                                                                                                                                                                                                                                     |                                                                                                                                                                        |
| Name of depositary institution<br>American Type Culture Collection                                                                                                                                                                                                                                                                                                                                                                             |                                                                                                                                                                        |
| Address of depositary institution (including postal code and country)<br>12301 Parklawn Drive<br>Rockville, Maryland 20852<br>UNITED STATES OF AMERICA                                                                                                                                                                                                                                                                                         |                                                                                                                                                                        |
| Date of deposit January 21, 1993                                                                                                                                                                                                                                                                                                                                                                                                               | Accession Number<br>A.T.C.C. 75399                                                                                                                                     |
| C. ADDITIONAL INDICATIONS (leave blank if not applicable) <span style="float: right;">This information is continued on an additional sheet <input type="checkbox"/></span>                                                                                                                                                                                                                                                                     |                                                                                                                                                                        |
| <p>"In respect of those designations in which a European patent is sought, a sample of the deposited microorganism will be made available until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample (Rule 23(4) EPC)."</p> |                                                                                                                                                                        |
| D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                        |
| E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                                                                        |
| The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")                                                                                                                                                                                                                                                                           |                                                                                                                                                                        |
| <div>For receiving Office use only</div> <div><input checked="" type="checkbox"/> This sheet was received with the international application</div> <div>Authorized officer </div>                                                                                                                                                                           | <div>For International Bureau use only</div> <div><input type="checkbox"/> This sheet was received by the International Bureau on:</div> <div>Authorized officer</div> |



## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

|                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                        |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| A. The indications made below relate to the microorganism referred to in the description<br>on page <u>31</u> , line <u>25</u>                                                                                                                                                                                                                                                                                                                 |                                                                                                                                                                        |
| B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input checked="" type="checkbox"/>                                                                                                                                                                                                                                                                                                                        |                                                                                                                                                                        |
| Name of depositary institution<br>American Type Culture Collection                                                                                                                                                                                                                                                                                                                                                                             |                                                                                                                                                                        |
| Address of depositary institution (including postal code and country)<br>12301 Parklawn Drive<br>Rockville, Maryland 20852<br>UNITED STATES OF AMERICA                                                                                                                                                                                                                                                                                         |                                                                                                                                                                        |
| Date of deposit<br>June 30, 1994                                                                                                                                                                                                                                                                                                                                                                                                               | Accession Number<br>A.T.C.C. 69341                                                                                                                                     |
| C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>                                                                                                                                                                                                                                                                                                        |                                                                                                                                                                        |
| <p>"In respect of those designations in which a European patent is sought, a sample of the deposited microorganism will be made available until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample (Rule 23(4) EPC)."</p> |                                                                                                                                                                        |
| D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                        |
| E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)<br>The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")                                                                                                                                                                                                  |                                                                                                                                                                        |
| <div>For receiving Office use only</div> <div><input checked="" type="checkbox"/> This sheet was received with the international application</div> <div><br/>Authorized officer</div>                                                                                                                                                                       | <div>For International Bureau use only</div> <div><input type="checkbox"/> This sheet was received by the International Bureau on:</div> <div>Authorized officer</div> |

## WE CLAIM:

1. A purified and isolated polynucleotide encoding a CviII polypeptide or a variant thereof possessing activity characteristic of CviII, said polynucleotide comprising a polynucleotide as set out in SEQ ID NO: 2.
2. The polynucleotide of claim 1 which is a DNA.
3. The DNA of claim 2 which is a viral genomic DNA sequence or a biological replica thereof.
4. The DNA of claim 2 which is a wholly or partially chemically synthesized DNA or biological replica thereof.
5. A purified isolated DNA encoding a polypeptide according to claim 1 by means of degenerate codons.
6. A vector comprising a DNA according to claim 2.
7. The vector of claim 6 which is the plasmid pCJH1.4 (ATCC Accession No. 69341).
8. A host cell stably transformed or transfected with a DNA according to claim 2 in a manner allowing the expression in said host cell of a CviII polypeptide or a variant thereof possessing a sequence specificity characteristic of CviII.
9. The host cell according to claim 8, wherein said host cell is *E. coli*.

10. A method for producing a CviJI polypeptide or a variant thereof possessing biological activity specific to CviJI, said method comprising the steps of:

- a) growing a transformed host cell containing a vector according to claim 6 in a suitable nutrient medium; and
- b) isolating the CviJI polypeptide or variant thereof from said host cell.

11. The method of claim 10 wherein said host cell is *E. coli*.

12. A recombinant CviJI polypeptide.

13. A polypeptide produced by the method of claim 10.

14. A method for restriction endonuclease digestion of DNA comprising the step of digesting DNA with a restriction endonuclease reagent under conditions wherein said DNA is cleaved at a dinucleotide sequence selected from the group consisting of PyGCPy, PuGCPy, PuGCPu, and wherein Pu = purine and Py = pyrimidine.

15. A method for restriction endonuclease digestion of DNA comprising the step of digesting DNA with a restriction endonuclease reagent under conditions wherein said DNA is digested at 11 of 16 possible dinucleotide sequences and wherein said dinucleotide sequences are selected from the group consisting of PuCGPu, PuCGPy, PyCGPy and PyCGPu, and wherein Pu = purine and Py = pyrimidine.

16. The method according to claim 14 wherein said restriction endonuclease reagent comprises CviJ I.

17. A restriction endonuclease reagent, said restriction endonuclease reagent comprising in combination, Taq I and Hpa II (CGase I), said reagent capable of digesting DNA at 11 of 16 possible dinucleotide sequences, said sequences selected from the group consisting of PuCGPu, PuCGPy, PyCGPy and PyCGPu, and wherein Pu = purine and Py = pyrimidine.

18. The method according to claim 15 wherein said restriction endonuclease reagent is selected from the group consisting of Aci I and CGase I.

19. The method according to claim 16 wherein said digestion of DNA is a partial digestion and wherein said digestion generates quasi-random fragments of DNA without apparent site preference as seen on a 1-2 wt. % agarose gel.

20. The method according to claim 18 wherein said digestion of DNA is a partial digestion and wherein said digestion generates quasi-random fragments of DNA without apparent site preference as seen on a 1-2 wt. % agarose gel.

21. The method according to claims 16 or 18 wherein said digestion is complete, and wherein said digestion generates DNA fragments from about 20 base pairs in length to about 200 base pairs in length and wherein said fragments have an average length of about 20 to about 60 nucleotides.

22. The method according to claims 19 or 20 wherein said quasi-random fragments are from about 100 basepairs to about 10,000 base pairs in length.

23. A method for shotgun cloning and sequencing DNA, comprising the steps of:

- a) partially digesting DNA according to claims 19 or 20;
- b) ligating said partially digested DNA into a linearized cloning vector thereby creating a recombinant vector;
- c) introducing said recombinant vector into a host cell;
- d) selecting said host cell for the presence of said recombinant vector;
- e) growing and amplifying said host cell containing said recombinant vector;
- f) isolating and purifying said recombinant vector from said grown and amplified host cells; and
- g) sequencing said DNA contained in said recombinant vector.

24. The method according to claim 23 wherein said restriction endonuclease reagent comprises CviI I.

25. The method according to claim 23 wherein said restriction endonuclease reagent comprises CGase I.

26. The method according to claim 23 wherein said quasi-random fragments are from about 100 base pairs to about 10,000 base pairs in length.

27. The method according to claim 23 wherein said quasi-random fragments are from about 500 bp to about 2,000 bp in length.

28. The method according to claim 23 wherein said cloning vector is selected from the group consisting of plasmids, phage, and cosmids.

29. The method according to claim 28 wherein said plasmid is pUC19.

30. The method according to claim 28 wherein said bacteriophage is  $\lambda$ .

31. The method according to claim 28 wherein said bacteriophage is M13.

32. The method according to claim 23 wherein said host cell is a bacteria.

33. The method according to claim 32 wherein said host cell is *E. coli*.

34. The method according to claim 23 wherein said sequencing is dideoxy sequencing.

35. A kit for the shotgun cloning of DNA, said kit comprising in association:

- a) a restriction endonuclease reagent, according to claims 16 or 18;
- b) a restriction endonuclease buffer;
- c) ligation buffer; and
- d) T4 DNA ligase.

36. The kit of claim 35 further comprising in association:

- e) competent host bacteria;
- f) chromatography matrix said matrix useful for the size selection of restriction endonuclease digested DNA;
- g) spin filters, said spin filters useful for the size selection of restriction endonuclease digested DNA;
- h) a cloning vector;
- i) positive control DNA useful in the monitoring of the efficiency of the said shotgun cloning; and
- j) molecular size marker DNA.

37. The kit according to claim 35 wherein said restriction endonuclease reagent comprises CviI I.

38. The kit according to claim 37 wherein said restriction endonuclease buffer endonuclease buffer is CviI I<sup>\*\*</sup> buffer.

39. The kit according to claim 35 wherein said restriction endonuclease reagent comprises CGase I.

40. The kit according to claim 39 wherein said restriction endonuclease buffer is CGase I buffer.

41. The kit according to claim 36 wherein said competent host bacteria is competent *E. coli* DH5 $\alpha$ F'.

42. The kit according to claim 36 wherein said chromatography matrix is Sephacryl-S500.

43. The kit according to claim 36 wherein said cloning vector is M13 mp18.

44. A method for labeling DNA, the method comprising the steps of:

- a) digesting an aliquot of template DNA with a restriction endonuclease reagent according to claim 21 and wherein said digestion generates sequence-specific DNA fragments;
- b) mixing an aliquot of undigested template DNA with said sequence-specific DNA fragments, denaturing said mixture of template DNA and sequence-specific DNA fragments thereby generating denatured template DNA and oligonucleotide primers.
- c) annealing said primers to said denatured undigested template DNA to form a DNA-primer complex;
- d) performing an extension reaction from said primers in said DNA-primer complex using a DNA polymerase in the presence of one or more nucleotide triphosphates and wherein at least one nucleotide triphosphate has a label.

45. The method according to claim 44 wherein said restriction endonuclease reagent comprises CviI I.

46. The method according to claim 44 wherein said restriction endonuclease reagent comprises CGase I.

47. The method according to claim 44 wherein said extension reaction is performed by a DNA polymerase.



48. The method according to claim 47 wherein said DNA polymerase is *Thermus flavus* DNA polymerase.

49. The method according to claim 44 wherein the one or more nucleotide triphosphates are selected from the group consisting of dATP, dCTP, dGTP, dUTP and dTTP.

50. The method according to claim 44 wherein said labeled nucleotide triphosphate is selected from the group consisting of  $^{32}\text{P}$ -labeled nucleotide triphosphates and  $^{33}\text{P}$ -labeled nucleotide triphosphates.

51. The method according to claim 44 wherein said labeled nucleotide triphosphate is selected from the group consisting of biotin-labeled nucleotide triphosphates, fluorescein-labeled nucleotide triphosphates, dinitrophenol-labeled nucleotide triphosphates, and digoxigenin-labeled nucleotide triphosphates.

52. A method for thermal cycle labeling DNA comprising the steps of:

- a) digesting an aliquot of template DNA with a restriction endonuclease reagent according to claim 21 and wherein said digestion generates sequence-specific DNA fragments;
- b) mixing an aliquot of undigested template DNA with said sequence-specific DNA fragments, denaturing said mixture of template DNA and said DNA fragments thereby generating denatured template DNA and oligonucleotide primers;
- c) annealing said primers to said denatured undigested template DNA to form a DNA-primer complex;
- d) performing an extension reaction from said primers in said DNA-primer complex using a DNA polymerase in the presence of one or more nucleotide triphosphates and wherein at least one nucleotide triphosphate has a label.
- e) heat-denaturing said labeled extension products;
- f) reannealing said excess primers with said template DNA and with said extension products;
- g) performing at least one additional extension reaction from said DNA-primer complex using a DNA polymerase.

53. The method according to claim 52 wherein said restriction endonuclease reagent comprises CviI I.

54. The method according to claim 52 wherein said restriction endonuclease comprises CGase I.

55. The method according to claim 52 wherein said DNA polymerase is a heat stable DNA polymerase.

56. The method according to claim 55 wherein said heat-stable DNA polymerase is *Thermus flavus* DNA polymerase or a functional fragment thereof.

57. The method according to claim 52 wherein said extension products also serve as templates.

58. The method according to claim 52 wherein said label is selected from the group consisting of fluorescein, dinitrophenol, biotin, and digoxigenin.

59. The method according to claim 52 wherein said label is selected from the group consisting of  $^{32}\text{P}$ ,  $^{33}\text{P}$ ,  $^3\text{H}$ ,  $^{14}\text{C}$ , and  $^{35}\text{S}$ .

60. The method according to claim 52 wherein steps e)-g) are repeated up to 20 times.

61. A kit for labeling DNA, said kit comprising in association:
- a) a restriction endonuclease reagent, according to claims 16 or 18;
  - b) a restriction endonuclease buffer; and
  - c) a labeling buffer.

62. The kit according to claim 61 wherein said restriction endonuclease reagent comprises CviI I.

63. The kit according to claim 62 wherein said restriction endonuclease buffer is CviI I\* restriction endonuclease buffer.

64. The kit according to claim 61 wherein said restriction endonuclease reagent is selected from the group consisting of CGase I and Aci I.

65. The kit according to claim 64 wherein said restriction endonuclease buffer is CGase I buffer.

66. The kit of claim 64 further comprising:

- d) a concentrated mixture of 1 or more nucleotide triphosphates;
- e) a DNA polymerase;
- f) control DNA, said control DNA being useful for monitoring the efficiency of labeling.

67. The kit according to claim 66 wherein said nucleotide mixture is an equimolar mixture of one or more nucleotides selected from the group consisting of dCTP, dTTP, dATP, and dGTP.

68. The kit according to claim 66 additionally comprising a labeled nucleotide selected from the group consisting of biotin-11-dUTP, digoxigenin-11-dUTP and fluorescein-11-dUTP.

69. The kit according to claim 66 additionally comprising a labeled nucleotide selected from the group consisting of  $^{32}\text{P}$ -labeled nucleotides,  $^{33}\text{P}$ -labeled nucleotides,  $^{14}\text{C}$ -labeled nucleotides,  $^{35}\text{S}$ -labeled nucleotides, and  $^3\text{H}$ -labeled nucleotides.

70. The kit according to claim 66 wherein said DNA polymerase is the Klenow fragment of DNA polymerase 1.

71. The kit according to claim 66 wherein said DNA polymerase is a thermostable DNA polymerase.

72. The kit according to claim 66 wherein said thermostable DNA polymerase is *Thermus flavus* DNA polymerase.

73. A method for universal thermal cycle labelling DNA comprising the steps of:

- a) mixing an aliquot of template DNA with a holo-enzyme of a thermostable DNA polymerase, whereby the polymerase provides endogenously purified DNA primers;
- b) denaturing said mixture of template DNA and said endogenous DNA primers;
- c) annealing said mixture of denatured template DNA and said endogenous DNA primers to form a DNA-primer complex;
- d) performing an extension reaction from said endogenous DNA primers in said DNA-primer complex using said DNA polymerase in the presence of one or more nucleotide triphosphates and wherein at least one nucleotide triphosphate has a label;
- e) heat-denaturing said labeled extension products;
- f) reannealing said endogenous primers with said template DNA and with said extension products;
- g) performing at least one additional extension reaction from said DNA-primer complex using a DNA polymerase.

74. The method according to Claim 73 wherein said heat-stable DNA polymerase is *Thermus flavus* DNA polymerase or a functional fragment thereof.

75. The method according to claim 73 wherein said extension products also serve as templates.

76. The method according to claim 73 wherein said label is selected from the group consisting of fluorescein, dinitrophenol, biotin, and digoxigenin.

77. The method according to claim 73 wherein said label is selected from the group consisting of  $^{32}\text{P}$ ,  $^{33}\text{P}$ ,  $^3\text{H}$ ,  $^{14}\text{C}$ , and  $^{35}\text{S}$ .

78. The method according to claim 73 wherein steps e)-g) are repeated up to 20 times.

79. A kit for labeling DNA, said kit comprising in association:

- a) a holo-enzyme of a thermostable DNA polymerase;
- and
- b) a DNA polymerase buffer.

80. The kit of claim 79 further comprising:

- c) a concentrated mixture of 1 or more nucleotide triphosphates;
- d) control DNA, said control DNA being useful for monitoring the efficiency of labeling.

81. The kit according to claim 80 wherein said nucleotide mixture is an equimolar mixture of one or more nucleotides selected from the group consisting of dCTP, dTTP, dATP, and dGTP.

82. The kit according to claim 80 additionally comprising a labeled nucleotide selected from the group consisting of biotin-11-dUTP, digoxigenin-11-dUTP and fluorescein-11-dUTP.

83. The kit according to claim 80 additionally comprising a labeled nucleotide selected from the group consisting of  $^{32}\text{P}$ -labeled nucleotides,  $^{33}\text{P}$ -labeled nucleotides,  $^{14}\text{C}$ -labeled nucleotides,  $^{35}\text{S}$ -labeled nucleotides, and  $^3\text{H}$ -labeled nucleotides.

84. The kit according to claim 80 wherein said thermostable DNA polymerase is *Thermus aquaticus* DNA polymerase.

85. The kit according to claim 80 wherein said thermostable DNA polymerase is *Thermus flavus* DNA polymerase.

86. A method for labeling of restriction-generated oligonucleotides, the method of comprising the steps of:

- a) digesting an aliquot of template DNA according to claim 21;
- b) heat denaturing said digested DNA thereby generating sequence-specific oligonucleotides; and
- c) labeling said sequence-specific oligonucleotides with a label capable of detection.

87. The method according to claim 86 wherein said restriction-generated oligonucleotides are labeled on the 5' end.

88. The method according to claim 86 wherein said restriction-generated oligonucleotides are labeled on the 3' end.

89. The method according to claim 86 wherein the label is radioactive.

90. The method according to claim 86 wherein the label is non-radioactive.



91. A method for anonymous primer cloning, the method comprising the steps of:

- a) digesting an aliquot of template DNA according to claim 21 thereby generating anonymous DNA fragments;
- b) digesting a plasmid cloning vector with a restriction endonuclease thereby creating a cloning site for insertion of said anonymous DNA fragments;
- c) ligating the anonymous DNA fragments of step a) into the cloning site of step b) thereby creating recombinant plasmids;
- d) transforming competent bacteria with the recombinant plasmids;
- e) selecting transformed colonies;
- f) purifying the recombinant plasmids from said transformed bacteria;
- g) digesting the recombinant plasmid with a restriction endonuclease said restriction endonuclease being capable of cutting said recombinant plasmid at a site, said site lying within the cloned anonymous DNA fragment;
- h) annealing one or more extension primers to the digested recombinant plasmid, said extension primers being complementary to plasmid sequences flanking the anonymous primer;
- i) extending the extension primer in a template-dependent fashion in the presence of one or more nucleotide triphosphates and a DNA polymerase; and
- j) denaturing the said hybridized extended primer.

92. The method according to claim 91 wherein said restriction endonuclease reagent comprises CviI I.

93. The method according to claim 91 wherein said restriction endonuclease reagent comprises CGase I.

94. The method according to claim 91 wherein said plasmid cloning vector is pFEM.

95. The method according to claim 94 wherein the restriction endonuclease of step b) is Eco RV.

96. The method according to claim 91 wherein said extension primer has a label capable of detection.

97. A kit for anonymous primer cloning comprising in association:

- a) a restriction endonuclease reagent, according to claims 16 or 18;
- b) a restriction endonuclease buffer;
- c) a cloning vector;
- d) competent bacteria;
- e) one or more extension primers said extension primers being complementary to plasmid sequences flanking said anonymous primers; and
- f) a DNA polymerase reagent.

98. The kit according to claim 97 wherein said restriction endonuclease reagent comprises CviJ I.

99. The kit according to claim 98 wherein said restriction endonuclease buffer is CviJ I\* buffer.

100. The kit according to claim 97 wherein said restriction endonuclease reagent is selected from the group consisting of CGase I and Aci I.

101. The kit according to claim 100 wherein said restriction endonuclease buffer is CGase I buffer.

102. The kit according to claim 97 wherein said cloning vector is pFEM.

The diagram illustrates the circular map of the p710 plasmid. Key features include the *lacZ'*/M.CviJI gene, two *EcoRI* sites, the *f1 ori*, the *Amp<sup>R</sup>* gene, and the *E. coli ori*.

Figure 2

10 20 30 40 50 60 70 80  
1 ATGCTTTTC GCACGTTAGA ACTATTCGCT GGTATAGCTG GTATTCACA TGGCTTCAGA GGTATATCTA CACGAGTTTC 80  
81 ATTCGTAGAA ATTAATGAAG ACCGACAAAA ATTCCTGAAA ACAAGTTTT CAGATGCATC TGTATTCAAT GACGTTACGA 160  
161 AATTTACCAA ATCGGACTTC CCAGAAGACA TAGACATGAT TACTGCGGGA TCCCGTGCCA CTGGGTTTAA TATTCGAGGT 240  
241 TCTAGAACTG GATTCGAACA CAAGGAATCC GGTCTCTTG CTGATGTTCT GCGAATCAGC GAAGAGTATA AACCTAAAAA 320  
321 AGTGTCTTCG GAAAGCTCCC ATATGTTGTC CCACACTTAC AATCTCGATG TCGTCTGAAA AAAGATGATG GAAATTTGTT 400  
401 ATTTCTGCAA GTGGTAACCT TGTGCGGCAT CAATTATAGG AGCCCATCAT CAACGCCACC GGTGGTTTTG TCTCGGATT 480  
481 CGAAAGATT ATCAACCAGA AGAAATAATT GTATCTGTGA ATGTACAAA GTTCGACTGG GAAATAATG AACCACTGG 560  
561 TCAAGTAGAC AATAAGAGTT ACCGAAATTC AACTCTTGTT CGTCTGGCAG GATATTCGCT GGTCCCGGAC CAGATCAGAT 640  
641 ATGCTTTTCA CCGTCTATTT ACAGGTGATT TGAGTCTAT GTGGAAACT ACCTTCACAC CTGGGACAAAT AATTGGCAGC 720  
721 GAACACAAAA AATGAAAAGG AACTTACGAT AAAGTCAATA ACCGGTATTA TGAGAACCAT GTGATTATT CTTTTCAAG 800  
801 AAAAGAAATT CATCGCGCTC CTCTAAATAT ATCCGTGAAA CCACGTGATA TTCCGGAGAA ACATAACGGA AAAACACTCG 880  
881 TAGATCGGGA AATGATCAAG AATATTTGGT GCACACCATG TGCTAGTTAT GGCACCTGCTA CTCTGGATG CAATGTTCTG 960  
961 ACAGACCTTC AGTCACATGC ACTTCTTACA CAAGTCAGGT TTTTATATAG GGTGTATGT GACCGACATT TGTCTGTAT 1040  
1041 ATGCTGTGCA TGGTTGATGG GGTATGACCA AGAATATCTT GGTATTTTGG TTCAATATGA TTAAATATT TTGATACAT 1120  
1121 AAATGGATAT AAGAAGAAAA CGTTTTACAA TAGAAGGGGC TAAACGTATA ATACTCGAAA AAAAGAGACT TGAAGAGAAA 1200  
1201 AAAAGAAATT CGGAAGAGAA AAAAGAAATT GCACCTTATG AAAACACACG AATTCGCGAA GAGAAAAAAA GAATTCGCGA 1280  
1281 AGAAGAAAAA CGATTTCGAC TTGAAGAGAA AAAACGAATT CGCGGAAGAG AAAAACAGAA AAAACACGAA 1360  
1361 TCGTGGAGAA GAAAAAAGA CTTCGACTTA TAGAAAAACA ACAGTAATGCG GAAGAGAAAA TTGCGTCTGG GAGAAAAATT 1440  
1441 AGAAAGAGGA TCTCTACAAA TGCAACAAAA CATGAAAGAG AATTTGTCAA AGTTATAAAT TCAATGTTCG TCGGACCCGC 1520  
1521 TACTTTTGTA TTCGTAGATA TAAAGGTA TAAATCCAGA GAAATCCACA ACCTTGTAAAG ATTCAGACAA TTACAAGGCA 1600  
1601 GTAAAGCGAA ATCCCGACCC GCGTATGTTG ATAGAGAATA TAACAAACCT AAAGCGGATA TAGACCGCGT AGACATAACC 1680  
1681 GGTAAAGATG TGGCATGGAT ATCCCATAAA GCATCTGAAG GATATCAACA ATATCTAAAA ATTTCTGGAA AGAACCTCAA 1760  
1761 GTTCACAGGA AAAGAATTCT AGAAGTTCT ATCTCTCAAG AGAAAGTAG TTAGTATGGC ACCGGTATCT AAAATATGGC 1840  
1841 CTGCTAATAA GACCGTATGG TCTCTATCA AGTCAAAATT GATTAATAAT CAAGCAATAT TCGGATTTGA TTACGGTAAG 1920  
1921 AAACCGAGGA GGGACAAATG AGACATCAA GGTCAAGGAC GACCAATTTT AACAAAAAGA GGTTCATAT TATATCTTAC 2000  
2001 ATTCACCTGG TTTAGCGCAT TAAATGGGCA CTGTCGTCAA CTGGAGAA TTACTGGGA AACATGAACC CGTTTTCTAT GTAAGAACAG 2080  
2081 AACCGAGTAG TAGCGGGAGA AGTATAACAA CTGTCGTCAA TGGTGTCACT TATAAAAAAT TAAGATTTCT TATACATCCA 2160  
2161 TACAACCTTG TTTCTTCAA AACACACGCT ATTAGTATAG ACCATTTTTCC CGAGAGACT TGTGTACCCG GTACTAAAAA 2240  
2241 ATGCTCAGCA TATTTGTCTA AAGATGCTCA TAGAAGCAGG TGCAAAACCT GACATCGTCA GTGTTGAGTA TACACCATTA 2320  
2321 CATCTACATG TGGTGATATT TGTATAACCG GTAAATACCT ATATATACAA TACGTATCCC CCTAAAAGCG CTTAGATTTT 2400  
2401 TTAGTTGAT ACTACTTTTG TATAAGACCT GTAAAGTACA AACATAAAGT TTCAGCTTTG CCTTCGAAAC AAGCAATTAC 2480  
2481 CCGCATGAGAA TAATATCCAT TATGGATTT TCTGCTAAT AAAACGATAT TTCTACAGA AGTTTCTATG ATTTGTTCCG 2560  
2561 AAATATTGAG ATCATCGTCA CGTTTTTCTT TACCGTATTT TACTTTCTGT ATCTGCTGCAC CAATAAAATC ATCTCTGTG 2640  
2641 AGTTCACTTG GCAATTGTGC CGTGACACCA AATCTCTAC AACCAACCTG ATGTCCATCC ATTTGCTAACA CTATCGGTA 2720  
2721 TCCATGTGTG GTGTGTACGA CCACACCGTT ATAACATATA CACGTGTAGT TGTCTCTAT ATCATATAA TCGAGAGCGG 2800  
2801 GTGAACTTG TTCAGATCTA TTATTAATCG GATCTGTACC AAGAAGAA TCTTCATAT TACAAATAAA ATCTCCGAT 2880  
2881 ATGTTCTGCA CACGAACAAC ATTCGTGAAA TTTCTGTGAT GACGAATCTC TACTCTGAA TCATTAGAGA CTTCGGAGTA 2960  
2961 TATAACATTA TAATTTGTTA TATGATTAT ACCTTTTATA TCAAAAAAT ACATATAAAC ACCATACAAA TATTAAGAT 3040  
3041 CGTTAGTATA TAATGGATAA CATTTGCAAT AGTATATTCA TCGCAGTAAA AAATGGCCAC GAAGCTTGT TGAAGATGAT 3120  
3121 GCTCTATGAA AGAGGTAGCA ATATCAATGA TGTTCGGAAT TCAAAAATG GAAATACACC ACTACATATT GCAGCTCATC 3200  
3201 ATGGTAATGA TGTGTGTTG AAGATGCTTA TTGACGACGG TGCAAAACCT GATATCACAG ATATTCTGAG AGGAACACCA 3280  
3281 CTTCATCTGT CGGTTTTGAA TGGCCATGAC ATATGTGTAC AGATGCTCGT AGAAGCAGGT GCAAAACCTTA GTATCATAA 3360  
3361 TAATTTGGGA TGGATACCGT TACATTACCG GCGTTTTAAT ACATACCGG GCTTTTAATG GTCATAGCAT GTGCGTCAA 3440  
3441 ATAATGTTGA CGTTATCAAT GATCGCGGTT GGACGGCGTT TTTCCGGATG TACACCACTT CATCGTCCGG TTTAATAG 3520  
3521 ACGCTTATTG ATGCGGGTGC AAATCTTGAC ATCAGAGATA ACTCTTGAC TCAATTGATG TACTGAGTGG GTGCCGTTAC 3600  
3601 ATTACGCGCG TTTTAATGGT AATGATGCGA TTTGAGGAT GCTCATTTGA GCAGGTGCAG ATATTGATAT ATCTAATATA 3680  
3681 TGTGTAATGGA CGGCGTTACA TTACGCGGCT CGAAATGGAC ACGATGTGTG TATAAAAAA CTCATCGAAG CAGGTGGTAA 3760  
3761 CATCAACCGC GTCAACAAAT CCGGGGATAC ACCATAGAT ATTTGCAAGT GTCATGACAT TGCAGTATGT GTGATCGTGA 3840  
3841 TAGTCAATTA GATCGTTTCG GAGCGGCGGT TCGCTCCGAG TGAGTTGTGT GTCATACCAC CAACGTCTCG TGCATTAGGT 4000  
4001 AGTGTGTTGG CAACGACGAT CGGCTTCAT GGGCGATCGG AAGCTGCAAA GATCACGCG CATCTTCTG TGGGTGCAAG 4080  
4081 GGATACTCTA CGACTATTT TTTCCAAAAA GGGTTGCAAT TTTACTTTAC CTACAGTATT ACCACTTCT TTTTTCCTA 4160  
4161 GTGAAAGTT ACCTATTTT CAATAGAGAC CATACGTACC TCCAAATTCA TTTACTTTAC CCGATGAAAC CGTTTGTGCG CAATTTTACC 4240  
4241 TGTATTCTGT TAAATTCAAC CTTTGAACCT CATCGCCATT AACAGACAGA GCGTATGAAC CGTTTGTGCG CAATTTTACC 4320  
4321 TTTCAAAACGA TAGTAACCCA TTGACCTCTA GGAATTTTAA CCGATCTTAT AAGTATCTGC TTACTTCCA GTCTTTTTC 4400  
4401 AAAAGCATAC AAGCATCTG TAAGGTTATC CCCAGAACCT GAAATTTGTA AGAACGACTG GAAATGAATA GGTGCAATTA 4480  
4481 GATCTGTATA CATATCACTT GGTTCGAAAT GAAATTCGTA GTCCCAATTA GGTACGTTCC ACCAAGTTTA ATACGGGGTC 4560  
4561 TTTCCACCGA GACCGGACAT TTGACGACGA GCCTTGTAGT AATGATATGA TGTGTTAAA TCTCTATCAC CATCGTTCCA 4640  
4641 TTTTCTCTG AACCGAAGAC CATGCATCGT TATACCTAGG GAAACCTGTA CTAAATCTT TATTTCAGGT GCGGCTCCGG 4720  
4721 GTGGATTAAC TCGAGATTTC TCAATCTAA AATATGATA CGATGTTCCA ACAGTAGAAC CACTGGGTGG TTTGCAAGGT 4800  
4801 GTGGAAGGG AAGGTAAAAA TTTAGGATAT TTCAATCAAC CAACACCTTG AGGTTTTACT TGAATACTTC TGGGAGATGT 4880  
4881 TGGTGGTTTC GTGGAAGGTG GTTTCGTTGA AGGTGGTTTC AGGTGGTTTC GTGGAAGGTG GTTTCGTTGA AGGTGGTTTC 4960  
4961 GTGGAAGGTG GTTTCGTTGA AGGTGGTTTC GTGGAAGGTG GTTTCGTTGA AGGTGGTTTC GTGGAAGGTG GTTTCGTTGA AGGTGGTTTC 5040  
5041 CCGAAGTGGG GCATGACCAT AATCGGTTAA ATTCCCGCAT TCACTTAATG ATGTAATCA TAAAGAACCG GGTGCGCAT 5120  
5121 GCATTCTTAT TGGTCTGTA GTATCAGATA TACATACGAA TCACTTAATG ATGTAATCA TAAAGAACCG GGTGCGCAT 5200  
5201 TGGCTTTTAC ATGACATTA TTGTAATATA ATATTATAT AAATTTAAAA AAACTAACGT CTATTTAAAA TTATGTAATA 5280  
5281 CGTATTATAT CAATGCATCA TCTTAATCAT TTCTAACCT AATAGCGTAG CGAATTC 5360  
5361

Figure 3

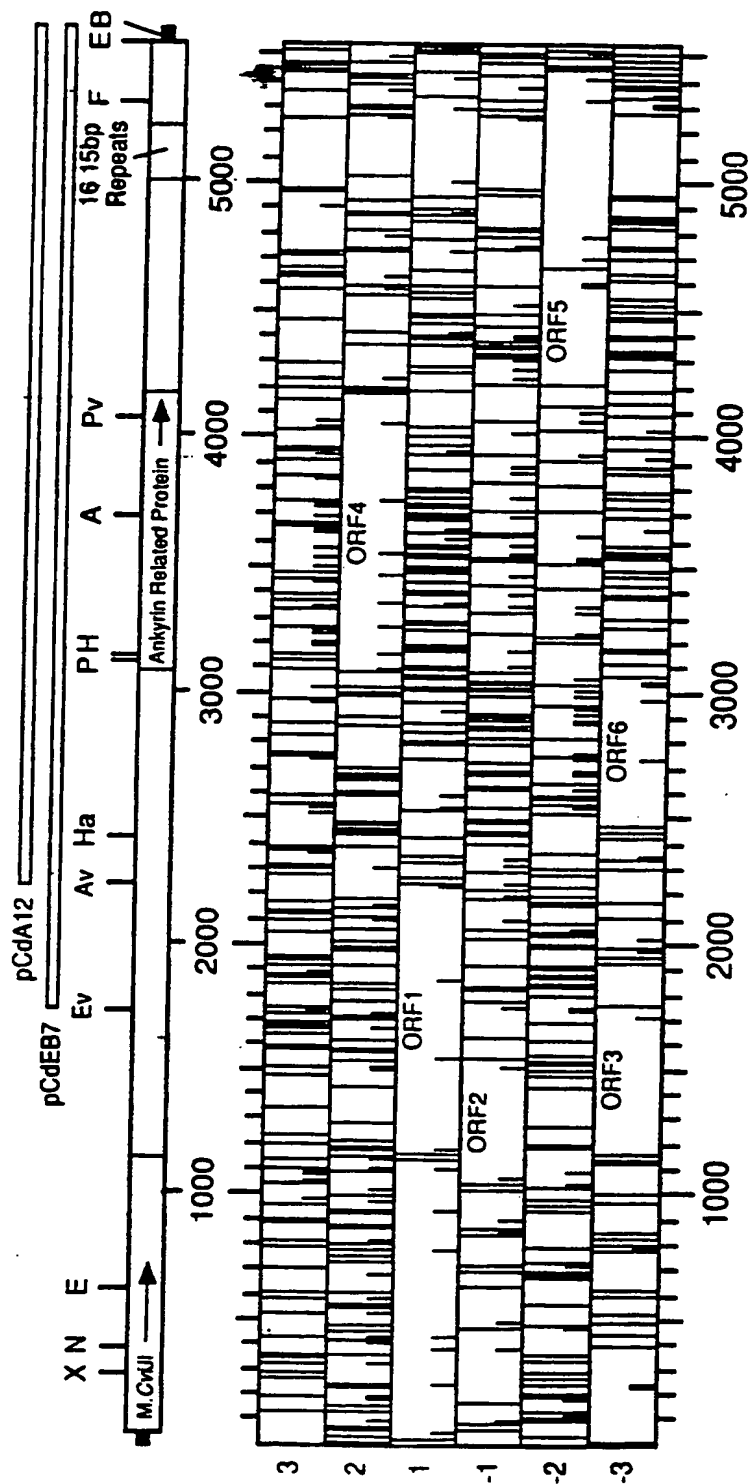


Figure 4

**M.CvJI**

1 CAA GAA TAT CTT GGT TAT TTG GTT CAA TAT GAT TAA AATATTTTGATACACTAA ATG GAT ATA  
Q E Y L G Y L V Q Y D . m d i

68 AGA AGA AAA CGT TTT ACA ATA GAA GGG GCT AAA CGT ATA ATA CTC GAA AAA AAG AGA CTT  
r r k r f t i e g a k r i i i e k k r i

128 GAA GAG AAA AAA AGA ATT GCG GAA GAG AAA AAA AGA ATT GCA CTT ATA GAA AAA CAA CGA  
e e k k r i a e e k k r i a i i e k q r

188 ATT GCG GAA GAG AAA AAA AGA ATT GCG GAA GAG AAA AAA CGA TTC GCA CTT GAA GAG AAA  
i a e e k k r i a e e k k r f a i e e k

**R.CvJI**

248 AAA CGA ATT GCG GAA GAA AAA AAA CGA ATC GCG GAA GAG AAA AAA CGA ATC GTG GAA GAG  
k r i a e e k k r i a e e k k r i y e e 3

308 AAA AAA AGA CTT GCA CTT ATA GAA AAA CAA CGA ATT GCG GAA GAG AAA ATT GCG TCG GGG  
k k r i a i i e k o r i a e e k i a s g 23

368 AGA AAA ATT AGA AAG AGG ATC TCT ACA AAT GCA ACA AAA CAT GAA AGA GAA TTT GTC AAA  
R K I R K R I S T N A T K H E R E F V K 43

428 GTT ATA AAT TCA ATG TTC GTC GGA CCC GCT ACT TTT GTA TTC GTA GAT ATA AAA GGT AAT  
V I N S H F V G P A T F V F V D I K G N 63

488 AAA TCC AGA GAA ATC CAC AAC GTT GTA AGA TTC AGA CAA TTA CAA GGC AGT AAA GCG AAA  
K S R E I H N V V R F R Q L O G S K A K 83

548 TCC CCG ACC GCG TAT GTT GAT AGA GAA TAT AAC AAA CCT AAA GCG GAT ATA GCA GCG GTA  
S P T A Y V D R E Y N K P K A D I A A V 103

608 GAC ATA ACC GGT AAA GAT GTG GCA TGG ATA TCC CAT AAA GCA TCT GAA GGA TAT CAA CAA  
D I T G K D V A W I S H K A S E G Y O O 123

668 TAT CTA AAA ATT TCT GGA AAG AAC CTC AAG TTC ACA GGA AAA GAA TTA GAA GAA GTT CTA  
Y L K I S G K N L K F T G K E L E E V L 143

728 TCG TTC AAG AGA AAA GTA GTT AGT ATG GCA CCG GTA TCT AAA ATA TGG CCT GCT AAT AAG  
S F K R K V V S H A P V S K I W P A N K 163

788 ACC GTA TGG TCT CCT ATC AAG TCA AAT TTG ATT AAA AAT CAA GCA ATA TTC GGA TTT GAT  
T V W S P I K S N L I K N O A I F G F D 183

848 TAC GGT AAG AAA CCA GGA AGG GAC AAT GTA GAC ATC ATA GGT CAA GGA CGA CCA ATT ATA  
Y G K K P G R D N V D I I G O G R P I I 203

908 ACA AAA AGA GGT TCC ATA TTA TAT CTT ACA TTC ACT GGT TTT AGC GCA TTA AAT GGG CAC  
T K R G S I L Y L T F T G F S A L N G H 223

968 TTG GAG AAT TTT ACT GGG AAA CAT GAA CCC GTT TTC TAT GTA AGA ACA GAA CCG AGT AGT  
L E N F T G K H E P V F Y V R T E R S S 243

1028 AGC GGG AGA AGT ATA ACA ACT GTC GTC AAT GGT GTC ACT TAT AAA AAT TTA AGA TTC TTT  
S G R S I T T V V N G V T Y K N L R F F 263

1088 ATA CAT CCA TAC AAC TTT GTT TCT TCA AAA ACA CAA CGT ATT ATG TAG GACCATTTCCTCGAG  
I H P Y N F V S S K T O R I M . 273

1152 AGACTTTGTTGACCGCGTACTAAAAATGGTCACGATATTTGCTAAAGATGCTCATAGAAGCAGGTGCAAACCTTGAC

Figure 5

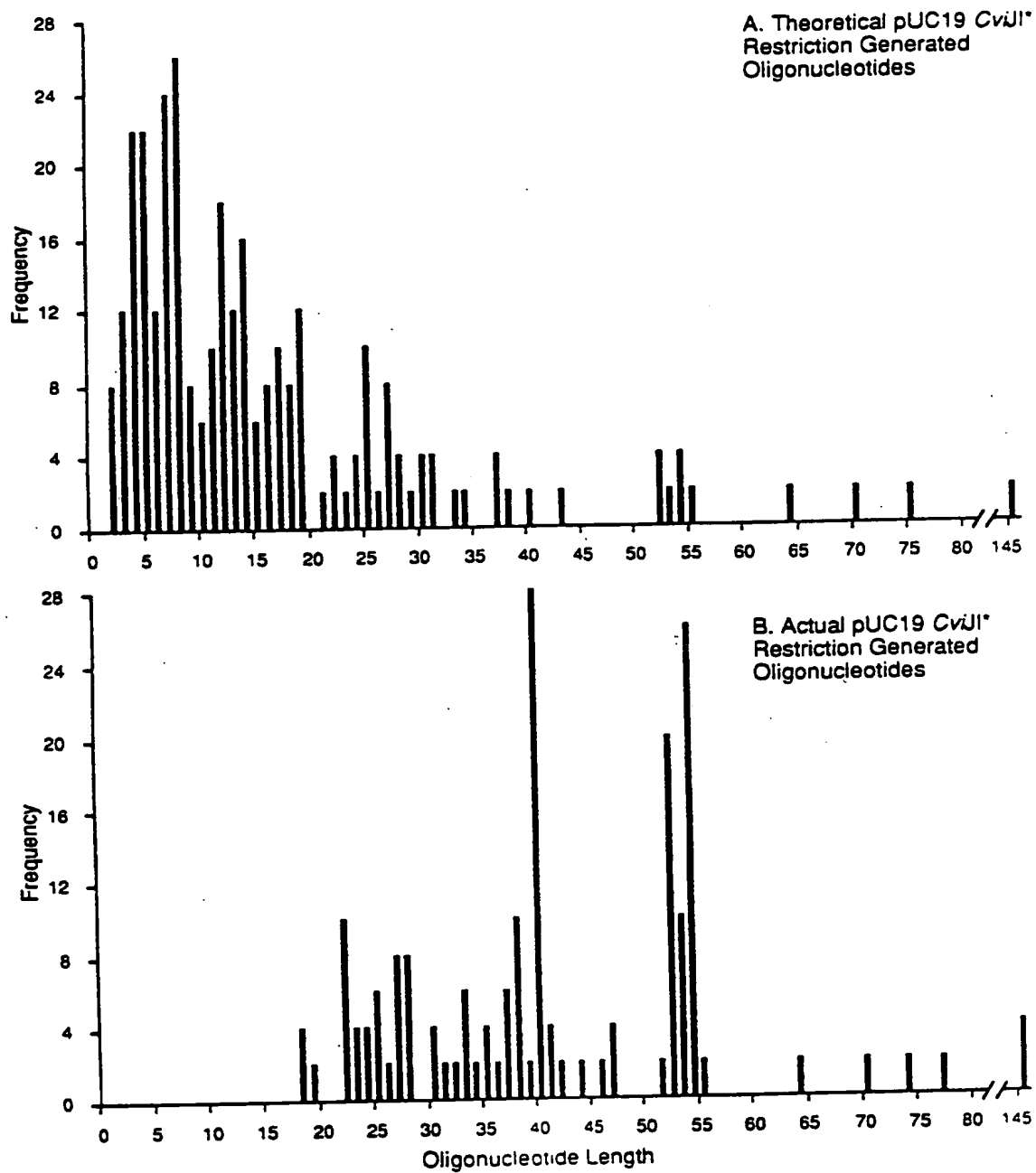




Figure 6

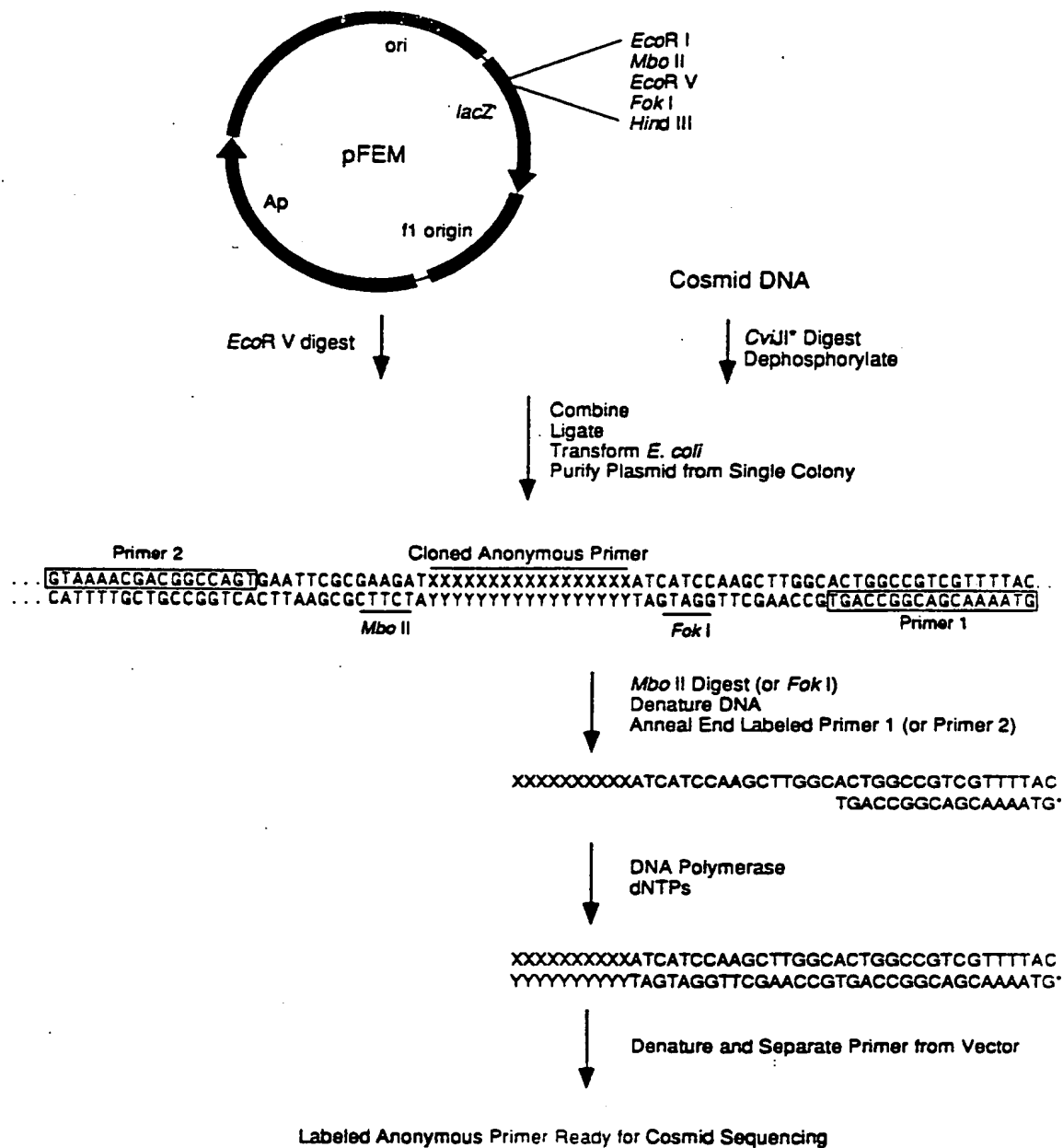


Figure 7

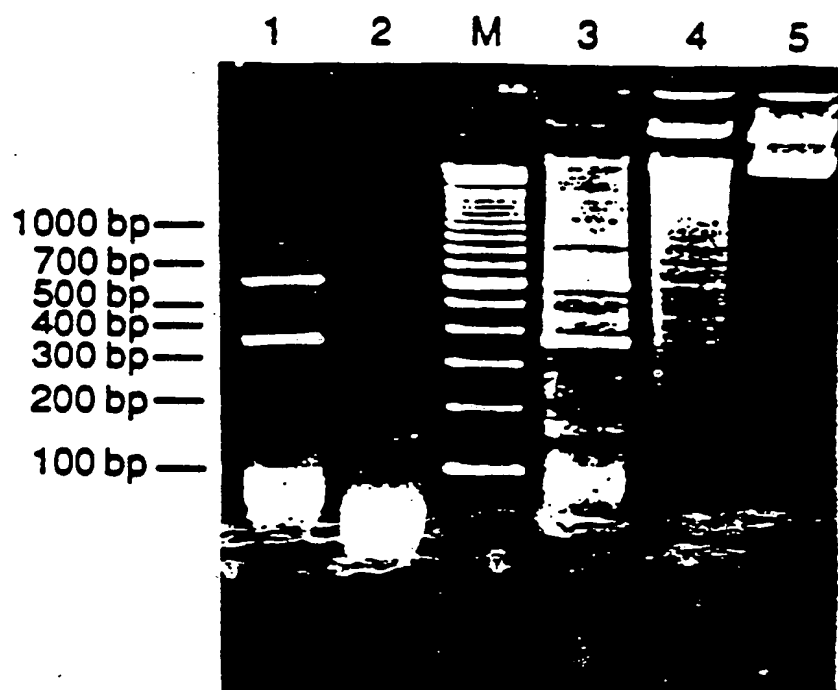


Figure 9

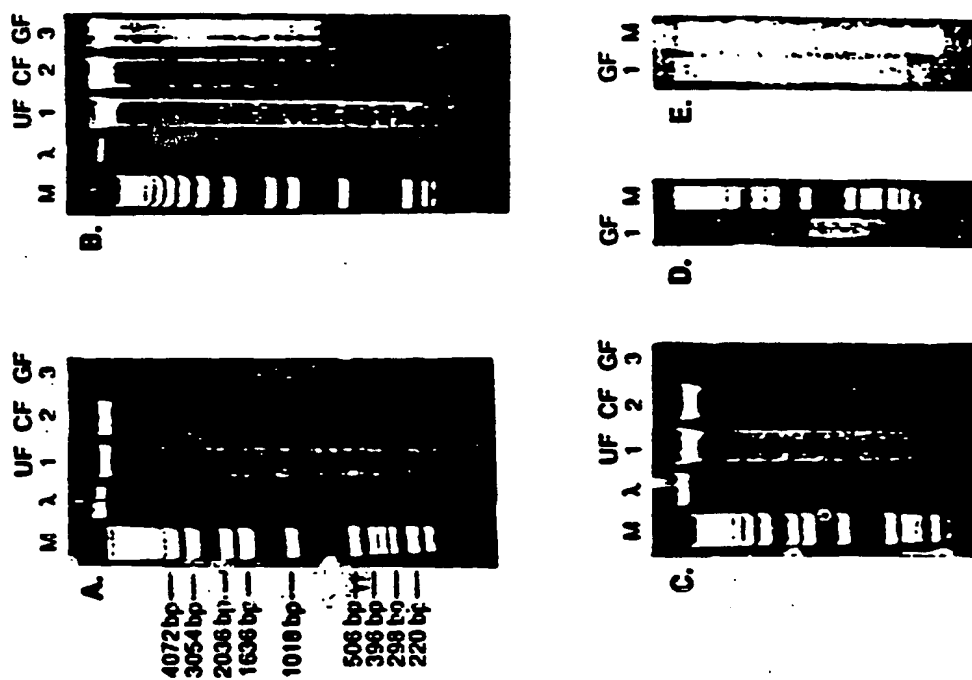


Figure 8

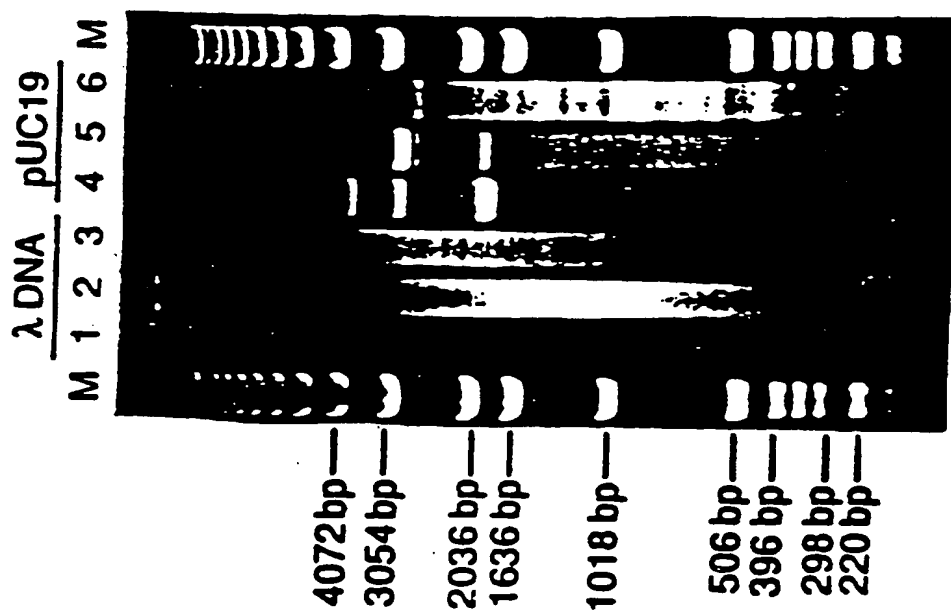


Figure 10

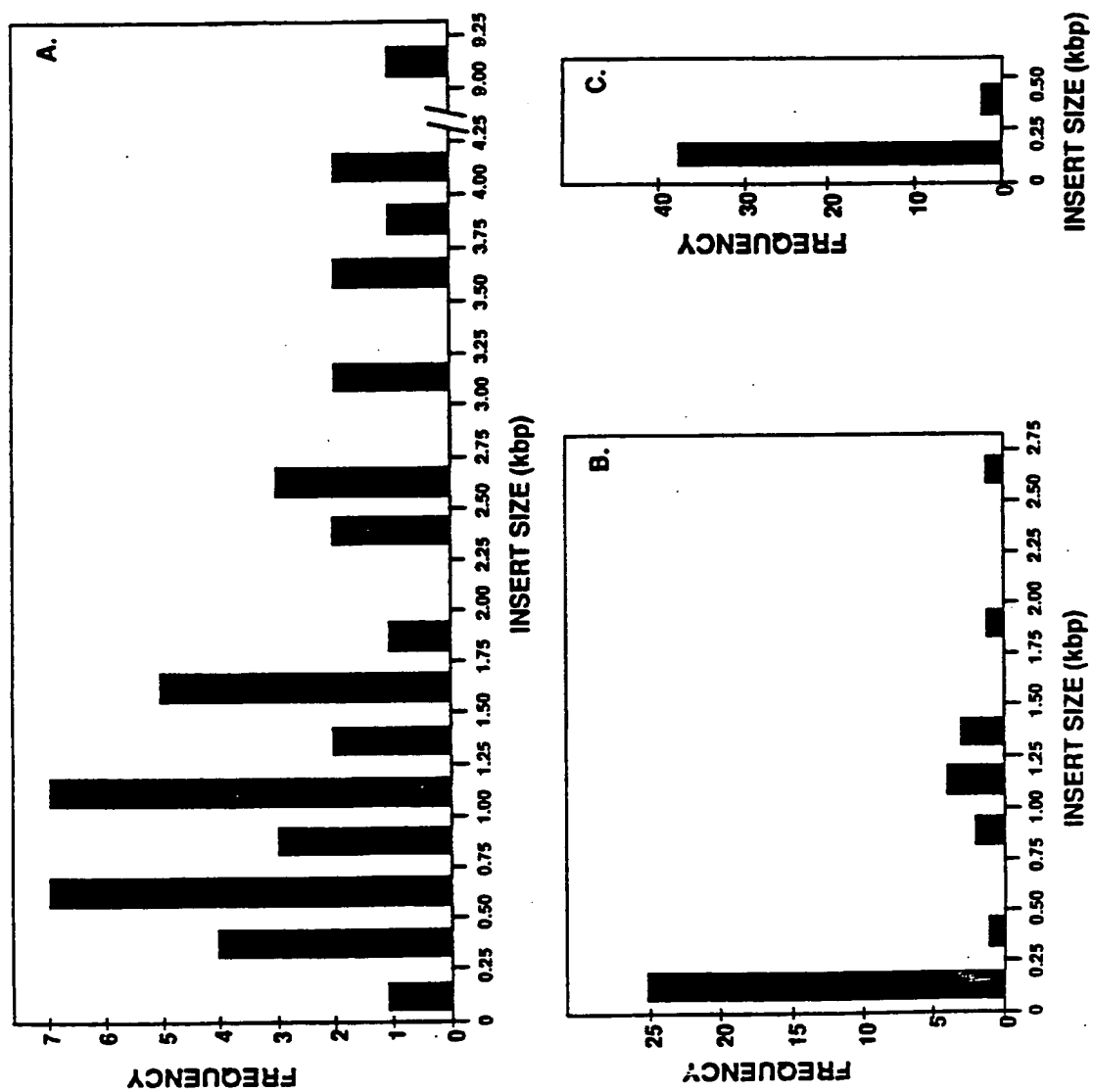


Figure 11

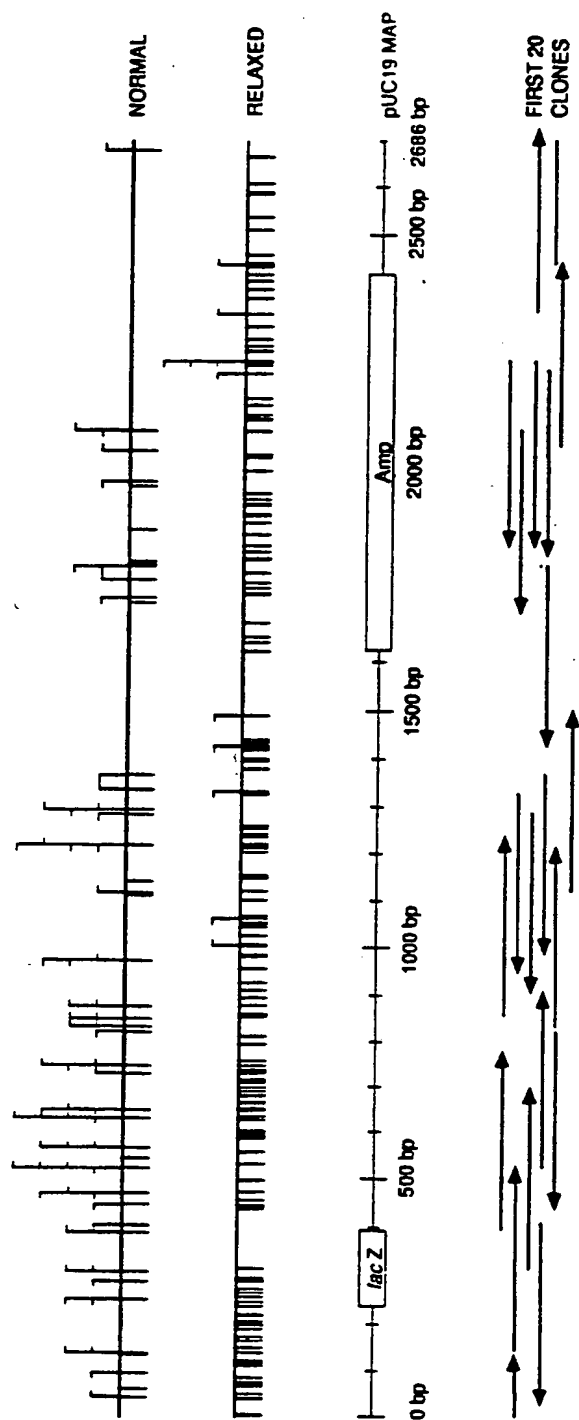
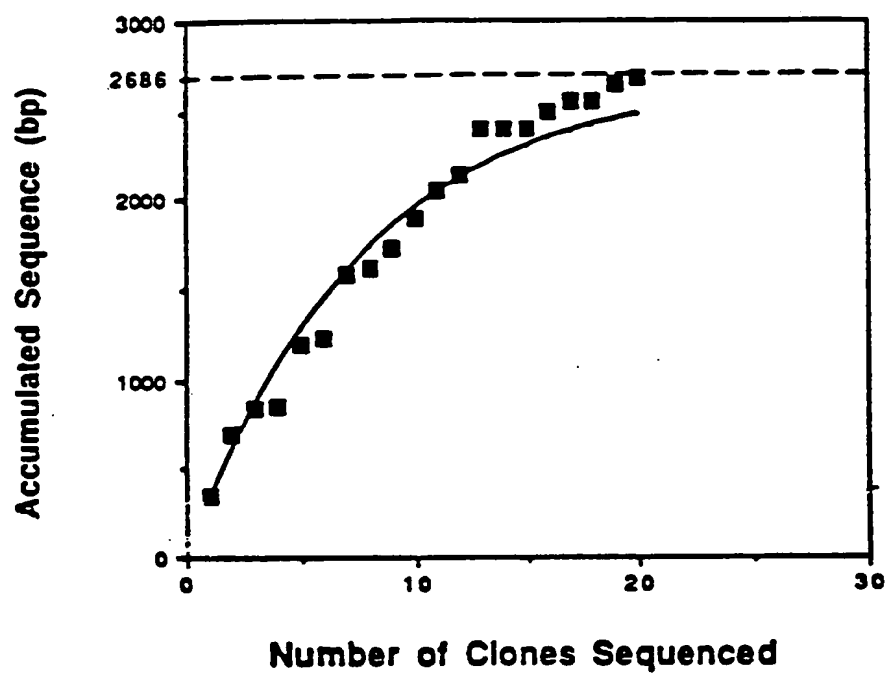


Figure 12



## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US94/03246**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(5) : C07H 21/02, 21/04; C07K 15/00; C12Q 1/68; C12P 19/34

US CL : 435/6, 91.1, 91.53, 172.3; 530/350; 536/23.72

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 91.1, 91.53, 172.3; 530/350; 536/23.72

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, MEDLINE, BIOSIS, CAS, WPI  
search terms: CviJI, cloned restriction**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                 | Relevant to claim No. |
|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| Y         | US, A, 5,053,330 (LUNNEN ET AL) 01 October 1991, column 2, lines 41-55.                                                                                                                                                            | 1-11                  |
| Y         | US, A, 5,075,232 (MORGAN) 24 December 1991, column 7, lines 32-43.                                                                                                                                                                 | 1-11                  |
| Y         | Nucleic Acids Research, Volume 20, No. 7, issued 1992, Y. Zhang et al, "A single amino acid change restores DNA cytosine methyltransferase activity in a cloned chlorella virus pseudogene", pages 1637-1642, especially Figure 2. | 1-11                  |
| Y         | Virology, Volume 176, issued 1990, S. L. Shields et al, "Cloning and Sequencing the Cytosine Methyltransferase Gene M. CviJI from Chlorella Virus IL-3A", pages 16-24, especially Figure 3.                                        | 1-11                  |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <input type="checkbox"/> Further documents are listed in the continuation of Box C.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | <input type="checkbox"/> See patent family annex.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| <p>* Special categories of cited documents:</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> | <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>"A" document member of the same patent family</p> |
| Date of the actual completion of the international search<br>29 JUNE 1994                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | Date of mailing of the international search report<br>JUL 13 1994                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| Name and mailing address of the ISA/US<br>Commissioner of Patents and Trademarks<br>Box PCT<br>Washington, D.C. 20231<br>Facsimile No. (703) 305-3230                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | Authorized officer<br>DAVID SCHREIBER <i>D. Schreiber</i><br>Telephone No. (703) 308-0196                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US94/03246

## BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

- I. Claims 1-11, drawn to a polynucleotide encoding CviJI, the vector carrying said polynucleotide, the transformed or transfected host carrying said vector, and method for producing a CviJI polypeptide from said host, classified in Class 536, subclass 23.72, for example.
- II. Claims 12 and 13, drawn to the recombinant CviJI polypeptide, classified in Class 530, subclass 350, for example.
- III. Claims 14, 16, 19, 21, and 22, drawn to a method for restriction endonuclease digestion using CviJI, classified in Class 435, subclass 6, for example.
- IV. Claims 14, 15, 17, 18, and 20-22, drawn to CGase I restriction endonuclease and a method for using it in restriction endonuclease digestion, classified in Class 435, subclass 6, for example.
- V. Claims 23, 24, 26-38, and 41-43, drawn to a method for shotgun cloning after partial digestion using CviJI, classified in Class 435, subclass 172.3.
- VI. Claims 23, 25-35, and 39-43, drawn to a method for shotgun cloning after partial digestion using CGase I, classified in Class 435, subclass 172.3.
- VII. Claims 44, 45, 47-53, and 55-63, drawn to a method of extension labeling of DNA and thermal cycle labeling using CviJI, classified in Class 435, subclass 91.1, for example.
- VIII. Claims 44, 46-52, 54-61, and 64-72, drawn to a method of extension labeling of DNA and thermal cycle labeling using CGase I, classified in Class 435, subclass 91.1, for example.
- IX. Claims 73-85, drawn to a universal thermal cycle labeling of DNA, classified in Class 435, subclass 91.1, for example.
- X. Claims 86-90, drawn to a method of end labeling after CviJI digestion, classified in Class 435, subclass 91.53.
- XI. Claims 86-90, drawn to a method of end labeling after CGase I digestion, classified in Class 435, subclass 91.53.
- XII. Claims 91, 92, and 94-99, drawn to a method for anonymous primer cloning after digestion with CviJI, classified in Class 435, subclass 172.3, for example.
- XIII. Claims 91, 93, 94-97, and 100-102, drawn to a method for anonymous primer cloning after digestion with CGase I, classified in Class 435, subclass 172.3, for example.

### Detailed Reasons for Lack of Unity

PCT Rule 13 recites the basic principle of unity of invention that an application should relate to only one invention or, if there is more than one invention, that applicant would have a right to include in a single application only those inventions which are so linked as to form a single general inventive concept. According to Rule 13, a group of inventions is linked to form a single inventive concept where there is a technical relationship among the inventions that involves at least one common or corresponding special technical feature that defines the contribution which each claimed invention, considered as a whole, makes over the prior art.

The thirteen inventions of this application consist of:

- 1) a polynucleotide encoding CviJI, the vector comprising it, the transformed host carrying the vector, and a method of making the protein using the vector,
- 2) the recombinant peptide CviJI,
- 3) a method for restriction endonuclease digestion using CviJI,
- 4) CGase I restriction endonuclease and a method for using it in restriction endonuclease digestion,
- 5) a method for shotgun cloning after partial digestion using CviJI,
- 6) a method for shotgun cloning after partial digestion using CGase I,
- 7) a method of extension labeling of DNA and thermal cycle labeling using CviJI,
- 8) a method of extension labeling of DNA and thermal cycle labeling using CGase I,
- 9) a universal thermal cycle labeling of DNA,
- 10) a method of end labeling after CviJI digestion,
- 11) a method of end labeling after CGase I digestion,
- 12) a method for anonymous primer cloning after digestion with CviJI, and
- 13) a method for anonymous primer cloning after digestion with CGase I.

The thirteen inventions are not linked by a special technical feature within the meaning of PCT Rule 13 for the following reasons: Those claims drawn to CviJI are not linked to those claims drawn to CGase I because there is no technical relationship among these inventions that involves at least one common or corresponding special technical feature.

The claims that involve the polynucleotide encoding CviJI, the vector containing it, the host carrying the vector, and methods of making recombinant protein are not linked to the recombinant protein because the protein and polynucleotide share a technical relationship that involves a corresponding technical feature that does not define the contribution which each claimed invention, considered as a whole, makes over the prior art because cloning and expression of polynucleotides to make recombinant polypeptides are well known in the art. Accordingly, such does not



# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US94/03246

constitute a special technical feature within the meaning of PCT Rule 13.2.

The methods for restriction endonuclease digestion, shotgun cloning and sequencing with CviJI, for extension and thermal cycle labeling with CviJI, for universal cycle labelling, for end labeling after CviJI digestion, and for anonymous primer cloning after CviJI digestion involve a corresponding technical feature, digestion with CviJI, that does not define the contribution which each claimed invention, considered as a whole, makes over the prior art because restriction endonuclease digestion, and shotgun cloning and sequencing, extension and thermal cycle labeling after rest, universal cycle labelling, end labeling, and anonymous primer cloning after restriction endonuclease digestion are well known in the art. In addition, CviJI is also known in the art. Accordingly, such does not constitute a special technical feature within the meaning of PCT Rule 13.2.

Similarly, the methods for restriction endonuclease digestion, shotgun cloning and sequencing with CGaseI, for extension and thermal cycle labeling with CGaseI, for universal cycle labelling, for end labeling after CGaseI digestion, and for anonymous primer cloning after CGaseI digestion involve a corresponding technical feature, digestion with CGaseI, that does not define the contribution which each claimed invention, considered as a whole, makes over the prior art because restriction endonuclease digestion, and shotgun cloning and sequencing, extension and thermal cycle labeling after rest, universal cycle labelling, end labeling, and anonymous primer cloning after restriction endonuclease digestion are well known in the art. Accordingly, such does not constitute a special technical feature within the meaning of PCT Rule 13.2.